Available online at www.sciencedirect.com

SCIENCE DIRECT

Chemometrics and Intelligent Laboratory Systems 75 (2005) 69–76

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

Standard error of prediction for multilinear PLS 2. Practical implementation in fluorescence spectroscopy

Rasmus Bro[a,*], Åsmund Rinnan[a], Nicolaas (Klaas) M. Faber[b]

[a] *Department of Dairy and Food Science, Food Technology, Chemometrics Group, Royal Veterinary and Agricultural University, Rolighedsvej 30, iii, 1958 Frederiksberg C, Denmark*

[b] *Chemometry Consultancy, Rubensstraat 7, 6717 VD Ede, The Netherlands*

Received 21 January 2004; received in revised form 11 April 2004; accepted 27 April 2004
Available online 7 August 2004

Abstract

In Part 1 of this series, a new simplified expression was derived for estimating sample-specific standard error of prediction in inverse multivariate regression. The focus was on the application of this expression in multilinear partial least squares (N-PLS) regression, but its scope is more general. In this paper, the expression is applied to a fluorescence spectroscopic calibration problem where N-PLS regression is appropriate. Guidelines are given for how to cope in practice with the main assumptions underlying the proposed methodology. The sample-specific uncertainty estimates yield coverage probabilities close to the stated nominal value. Similar results were obtained for standard (i.e., linear) PLS regression and principal component regression on data rearranged to ordinary two-way matrices. The two-way results highlight the generality of the proposed expression.

Keywords: Multiway calibration; Partial least squares; Standard error of prediction; Fluorescence spectroscopy; Noise addition

1. Introduction

It is considered good analytical practice to report a result together with an estimate of its uncertainty. For example, when fitting a line through scattering x–y points, it is desirable to construct the familiar confidence and prediction bands using well-known expressions from basic statistics. Interestingly, calibration methodology differs greatly when moving to more complex data structures, i.e., multivariate and multiway data. Beyond univariate calibration, the only generally accepted approach to prediction uncertainty is to use an overall measure such as root mean square error of prediction (RMSEP), hence for any prediction the uncertainty is set to a constant value [1]. Clearly, a negative aspect of this global uncertainty estimate is that it does not yield realistic prediction intervals. However, the required sample-specific prediction uncertainty estimates are available for latent variable (LV) methods in general, see Ref. [2] for a review of various proposals. Very recently, a simple and user-friendly expression for standard error of prediction (SEP) has been derived and tested on multiway models using Monte Carlo simulations [3]. It performed well for standard (i.e., linear) partial least squares (PLS) regression applied to near infrared (NIR) data sets [4]. The purpose of the current work is to:

- demonstrate the practical utility of our simple expression on fluorescence data, and
- identify the conditions that are critical for reliable use of the proposed methodology.

2. Theory

Since a detailed discussion of the proposed expression is given in Part 1 [3], we restrict ourselves to pointing out the relationship with variations of RMSEP, as well as another expression for sample-specific SEP. These relationships may lead to a better understanding of the properties of our proposed expression. To simplify the presentation, we have found it convenient to slightly adapt the earlier notation.

* Corresponding author.
E-mail address: rb@kvl.dk (R. Bro).

0169-7439/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2004.04.014

2.1. Prediction uncertainty on the global set level

Current practice is to characterise (multivariate or multiway) prediction uncertainty on the set level. An RMSEP-value is calculated as the root mean squared difference between predictions and reference values. It is important to stress that this procedure is only sound when the noise in the reference values is negligible compared with the true prediction uncertainty. The reason for this is that prediction errors are defined with respect to the true quantities, rather than noisy reference values. Consider the ideal situation where one has the perfect model and noisy reference values—a mental experiment. Of course, this limit is not practical, but adding noise to the reference values as described by DiFoggio [5] and Coates [6] can always approach it to some extent. Clearly, the predictions should be perfect and the only contribution to RMSEP would originate from the measurement error in the reference values. In this extreme case, RMSEP would just estimate the standard deviation (square root of the variance) of the measurement error of the reference value—it would not relate to the true prediction uncertainty at all! Thus, in general, the presence of this spurious error component leads to a so-called *apparent* RMSEP [5]:

$$\mathrm{RMSEP}_{\mathrm{app}} = \left[\frac{1}{I}\sum_{i=1}^{I}(\hat{y}_i - y_{\mathrm{ref},i})^2\right]^{1/2} \quad (1)$$

where I denotes the number of samples in the test set, $\hat{y}_i$ is the prediction of property y for sample i and $y_{\mathrm{ref},i}$ is the associated reference value. A simple but effective correction for the spurious error component leads to a so-called *corrected* RMSEP [5,7]:

$$\mathrm{RMSEP}_{\mathrm{cor}} = [\mathrm{MSEP}_{\mathrm{app}} - V_{\Delta y}]^{1/2} \quad (2)$$

where $V_{\Delta y}$ is an estimate for the measurement error variance associated with the reference method. This correction has been used successfully for NIR [8] and Raman [9] applications. If knowledge on $V_{\Delta y}$ is lacking, it can be set to zero and the more pessimistic apparent RMSEP is then obtained.

2.2. Prediction uncertainty on the individual sample level

Characterising prediction uncertainty on the set level is the best way to answer important questions like "how good is my calibration?" It is therefore logical, for example, to monitor changes in the (set level) RMSEP when optimising a calibration model (spectral pre-treatment, factor selection, etc.). However, as noted before, this procedure does not lead to sample-specific prediction intervals with good coverage probability. The American Society for Testing and Materials (ASTM) has recognised the need for a sample-specific SEP [10] and recommends using the expression originally proposed by Höskuldsson [11]:

$$\mathrm{SEP}_{\mathrm{app},i} = [(1 + h_i)\mathrm{MSEC}_{\mathrm{app}}]^{1/2} \quad (3)$$

where h_i symbolises the leverage for sample i and MSEC stands for the mean square error of calibration. This expression is implemented in certain commercial software [12].

The leverage is related to the distance of a sample to the mean of the calibration set data. The calculation of MSEC is similar to the calculation of the apparent (set level) RMSEP, i.e., Eq. (1), but now one has to account for the degrees of freedom of the calibration model. Because MSEC is explicitly based on reference values, Eq. (3) leads to an apparent sample-specific SEP when the reference method is imprecise. In other words, Eq. (3) is the sample-specific analogue of Eq. (1). Clearly, the correction in Eq. (2) can also be applied on the sample level, leading to our proposal [3]

$$\mathrm{SEP}_{\mathrm{cor},i} = [(1 + h_i)\mathrm{MSEC}_{\mathrm{app}} - V_{\Delta y}]^{1/2} \quad (4)$$

which was derived in Part 1 based on an approximation of a local linearization approach.

Notice that only the leverage reflects the individual differences. As leverages are simple to calculate this expression is highly operational. The quality of the expression is crucially dependent on the quality of the estimate of $\mathrm{MSEC}_{\mathrm{app}}$. Ensuring that the calibration model is sound, robust and representative is therefore of significant importance to be able to rely on the use of Eq. (4). This is in fact similar to the situation in univariate regression. For example, a model where the MSEC changes significantly by the in- or exclusion of one sample cannot be considered to provide a robust hence reliable estimate of MSEC. Likewise, if the cross- or test set-validated mean square error differs markedly from MSEC, then most likely the estimate cannot be considered to be appropriate. Some practical guidelines on how the adequateness can be assessed are provided in Experimental.

2.3. Validation of proposed sample-specific uncertainty estimate

Comparing the coverage probabilities of the resulting prediction intervals with the nominal value would validate Eq. (4). Unfortunately, this requires error-free reference values. However, the direct relationship between Eqs. (3) and (4), ensures that an equivalent test follows from the studentised *apparent* prediction residuals,

$$t_i = \frac{\hat{y}_i - y_{\mathrm{ref},i}}{\mathrm{SEP}_{\mathrm{app},i}} \qquad i = 1, \cdots, I \quad (5)$$

These should be approximately distributed as Students t with degrees of freedom (f) associated with the MSEC estimate [3]. In particular, the standard deviation should be close to $\sqrt{f/(f-2)}$. As $\mathrm{SEP}_{\mathrm{app}}$ does not rely on the actual reference measurements in the validation set, the measurement errors in these are of no consequence for the evaluation.

2.4. Checking validity of the regression model

For a specific regression model it is possible to verify whether the approach for estimating the sample-specific SEP (Eq. (4)) is appropriate. As described in Part 1, the formula is based on a local linearization and the validity of this first-order approximation can be tested. In this paper a noise-addition approach will be used for this purpose. From a principal component analysis (PCA) model of the predictor (spectral) data, the noise level in **X** can be determined. Adding different multiples of this level of random Gaussian noise to **X** leads to different realizations of the regression vector, and associated predictions of **y**. If the first-order approximation is valid, the standard error in the regression vector and hence in the predictions should increase linearly with the noise added. The presence of linearity can then be verified formally or visually in simple manners as will be shown in the experimental part. Test set predictions are used for assessing linearity as fitted values can led to spurious results when performed in an unsupervised fashion (fixing the number of components). Noise is added to both the calibration and the test data.

3. Experimental

3.1. Generation of the data

The data set used in this paper is part of a larger data set prepared for the study of several topics in fluorescence spectroscopy. The part not used here is characterised by specific artificially induced problems (e.g. co-varying components). The selected analytes have very similar excitation and emission spectra (see Fig. 1). Consequently, the calibration problem is fairly difficult.

The fluorescence spectra of 131 samples were recorded. Five different analytes were used: catechol (Sigma, approx. 99%), hydroquinone (Riedel-deHaën, min. 99.5%), indole (Riedel-deHaën, min. 99%), L-tryptophan (Merck, min. 99%) and/or DL-tyrosine (Sigma, min. 98%). All samples were mixtures of 2 to 4 of these fluorophores. The concentration ranges of the fluorophores in the samples are stated in Table 1.

The samples were prepared through several dilution steps with deionised water. First, a small amount of each analyte was weighted and transferred to a container. It was further diluted into standard strength, before they were mixed and diluted to the desired concentrations. The prepared samples were then measured by a Varian Eclipse Fluorescence Spectrometer. The settings for the instrument were: slit widths 5 nm (for both excitation and emission), emission wavelengths 230–500 nm (recorded every 2 nm) and excitation wavelengths 230–320 (recorded every 5 nm), scan rate 1920 nm/min and a PMT (photo multiplier tube) detector voltage of 600 V. The sample was excited

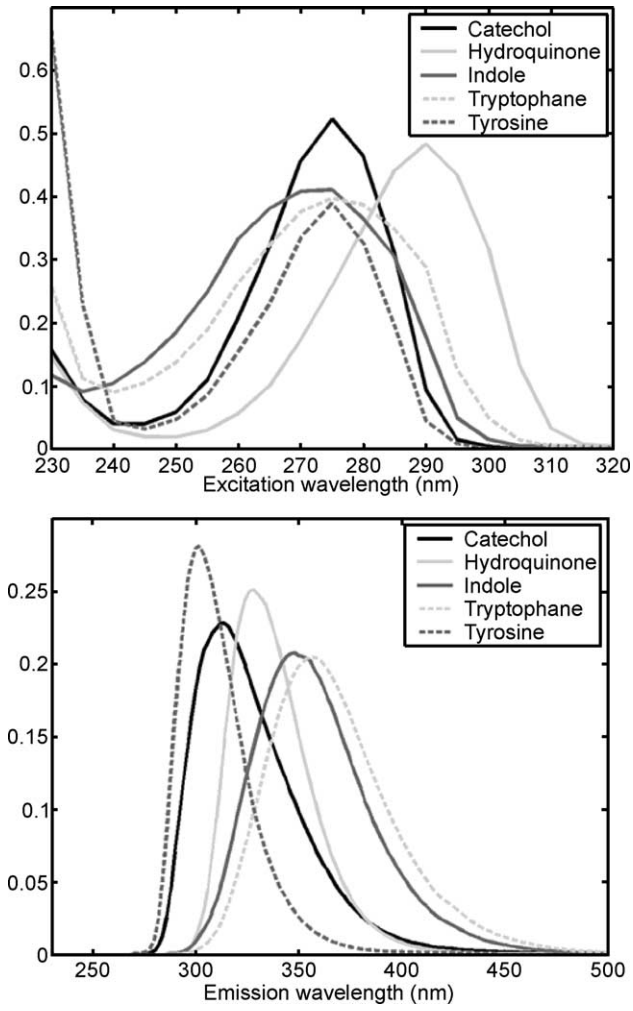

Fig. 1. The pure excitation and emission spectra for the five fluorophores used in the data set.

with lowest energy (highest excitation wavelength) first and then up to the highest energy excitation. Every sample was left in the instrument for a total of five replicate scans. The total recording time for one sample was approximately 15 min.

Everyday a standard was run before and after analysis in order to ensure that there was no drift in the instrument.

3.2. Outlier detection and wavelength selection during calibration

Eq. (4) only accounts for the effect of random noise in the data as well as the model [3]. Consequently, it is crucial for the analyses performed here, that the calibration model is well behaving. The term well behaving is difficult to quantify, but it means that the model is primarily reflecting the systematic variation foreseen in new samples. This can be quantified by assessing how close the model fit is to the (cross-) validated fit or by influence analysis. In line with this requirement, it is of importance to remove abnormal and

extreme samples as well as irrelevant variables. Such outliers and irrelevant variables deteriorate any sound statistical evaluation of the model and lead to misleading statistics. Outlying behaviour can come from: pollution in the sample, irregular behaviour of the instrument, incorrect sample preparation, extremely high or low concentration of an analyte, etc.

One method for outlier detection is based on initial PLS analyses, and the visual inspection of the T vs. U-score plots. These plots describe the relationship between X and Y. For a good prediction model with relevant variables, these plots should be approximately straight lines indicating a predictive relation between X and Y. If a sample diverges significantly from this line compared to the other samples, it is an outlier—its relationship between X and Y is different from the rest.

Prior to analysis, part of the recorded data was removed in order to avoid any scattering effects that are present in fluorescence spectroscopy [13]. The emission wavelength ranges 230–296 and 422–500 nm were removed, together with the excitation wavelength ranges 230–240 and 300–320 nm. This reduced the landscapes from 136 × 19 (emission × excitation) down to 62 × 10. Further, only the first replicate measurement of each sample was used in this analysis.

3.3. Calculation of the measurement error in the reference values

The samples were prepared in several dilution steps. Consequently, there are several uncertainties to be accounted for upon estimating the final uncertainty of the reference values. In order to calculate the uncertainty induced by the different dilution steps, the following error propagation formula has been used:

$$\sigma(y) = \sqrt{\sum_{q=1}^{Q}\left(\frac{\partial y}{\partial x_q}\right)^2 \sigma(x_q)^2} \tag{6}$$

where 'σ' denotes the standard deviation in the associated variable, '∂' symbolises a partial derivative, Q is the total number of parameters with uncertainties and x_q ($q = 1,\ldots, Q$) is a parameter of y.

The steps involved in making the solution were as follows: an amount of solid was weighed and transferred to a 250 ml container, 10 ml of this was taken out using a pipette, and transferred to a 100 ml container and 0.5 to 2 ml of this was taken out using an adjustable pipette and transferred to a 10 ml container.

Table 1
Concentration ranges for the analytes

Analyte	Concentration in 10^{-6} M
Catechol	0–87.0
Hydroquinone	0–22.5
Indole	0–5.46
Tryptophan	0–7.44
Tyrosine	0–12.14

Table 2
Measurement uncertainties in the equipment used in the preparation of the samples

Parameter	Amount	Uncertainty
Weight	0–1 g	0.0001 g
$\text{Volume}_{C,1}$	250 ml	0.23 ml
$\text{Volume}_{C,2}$	100 ml	0.1 ml
$\text{Volume}_{C,3}$	10 ml	0.038 ml
$\text{Volume}_{P,1}$	10 ml	0.03 ml
$\text{Volume}_{P,0.5ml}$	0.5 ml	0.0075 ml
$\text{Volume}_{P,1ml}$	1 ml	0.008 ml
$\text{Volume}_{P,2.5ml}$	2.5 ml	0.015 ml

An example of the uncertainty for one of these steps is as follows:

$$\sigma(V_{\text{new}}) = \sqrt{\left(\frac{\partial V_{\text{new}}}{\partial V_{\text{old}}}\right)^2 \sigma(V_{\text{old}})^2 + \left(\frac{\partial V_{\text{new}}}{\partial \text{vol}_P}\right)^2 \sigma(\text{vol}_P)^2 + \left(\frac{\partial V_{\text{new}}}{\partial \text{vol}_C}\right)^2 \sigma(\text{vol}_C)^2} \Rightarrow$$
$$\sigma(V_{\text{new}}) = \sqrt{\left(\frac{\text{vol}_P \sigma(V_{\text{old}})}{\text{vol}_C}\right)^2 + \left(\frac{V_{\text{old}} \sigma(\text{vol}_P)}{\text{vol}_C}\right)^2 + \left(\frac{V_{\text{old}} \text{vol}_P \sigma(\text{vol}_C)}{(\text{vol}_C)^2}\right)^2} \tag{7}$$

where V_{new} and V_{old} denote the new and old concentration, respectively, vol_P is the volume of the pipette used to transfer the analyte, and vol_C is the volume of the new container. Estimates of the measurement uncertainties for the different steps are most often given on the measuring equipment itself (see Table 2).

Each analyte was present at four levels of concentration, and can thus be coded from 1 to 4. Since the measurement error of the weight was constant over its range, and the relative uncertainty of the last pipette ($V_{\text{P,0.5 ml}}$, $V_{\text{P,1 ml}}$, and $V_{\text{P,2.5 ml}}$) increased with decreasing volume, the uncertainties varied among the analytes, as well as for the different concentrations (see Table 3). The relative uncertainties vary from 0.8% for the highest concentration of catechol, hydroquinone and tyrosine, to 2.1% for the lowest concentration of indole. In the remainder of this paper the lowest reference uncertainty—0.8%—is used as a lower bound of the uncertainty. This will give the most pessimistic results for the confidence limits (see Eq. (4)). In this way, the presentation of unrealistically good results is avoided.

Table 3
Uncertainty in the reference value for the different analytes with varying concentration

Analyte	Uncertainty in %			
	Relative concentration			
	1	2	3	4
Catechol	1.6	0.9	0.9	0.8
Hydroquinone	1.6	1.0	1.0	0.8
Indole	2.1	1.6	1.6	1.5
Tryptophane	1.7	1.1	1.1	0.9
Tyrosine	1.6	1.0	1.0	0.8

4. Results and discussion

Calibration models were constructed using both N-PLS as well as standard PLS and PCR. Only the results of N-PLS are shown in the following, but similar results were obtained for the two-way calibration models.

4.1. Constructing the model

The data set was divided into a calibration set of 35 samples and a validation set of 86 samples.

Initial PLS models were made for each analyte in order to check for gross outliers. Two *T* vs. *U*-score plots—showing the relationship between X and Y—showed one outlier each, see Fig. 2, hence the total number of samples were reduced to 129. Other traditional outlier diagnostics were also investigated, but no additional *gross* outliers were identified. By close inspection of the experimental setup and the two spectra in question, it was clear that the two samples were prepared wrongly. For the first outlier, the amount of catechol in the sample was less than it should be according to the experimental setup. For the second outlier the hydroquinone concentration was higher than planned. So, not only do these outliers make sense from a statistical point of view, but it is also possible to backtrack and investigate the actual reason for the outlying behaviour.

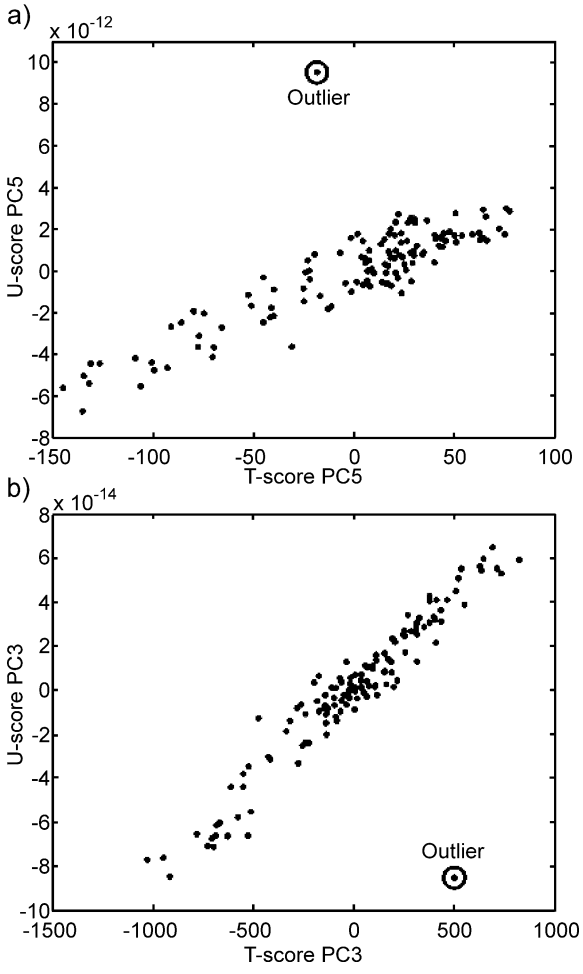

Fig. 2. *T* vs. *U* plots from PLS of (a) Catechol—PC5, and (b) Hydroquinone—PC3, both showing one clear outlier.

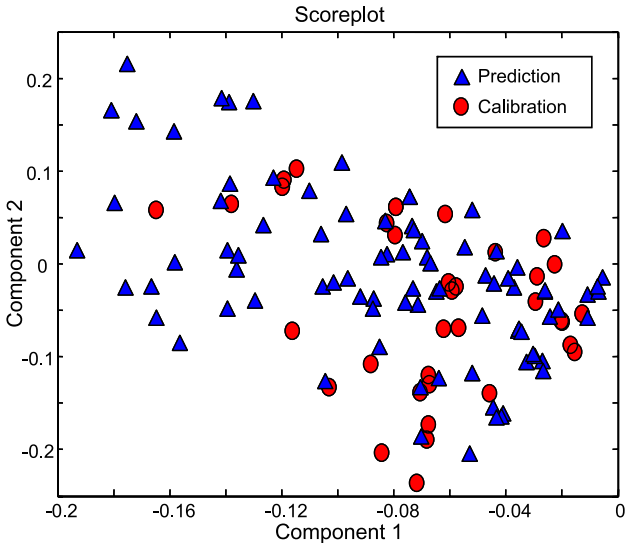

Fig. 3. Second (14% variance explained) versus first (68%) principal component using calibration and prediction spectral data. The plot shows how the prediction samples extend beyond the space of the calibration samples.

Eight additional samples were excluded during specific model building as potential outliers. These samples were not extreme outliers as the above-mentioned. For most, reasonable explanations for the outlying behaviour were possible to find but not for all. However, in order for the distributional properties of the uncertainties to be meaningful, even moderate outliers are necessary to remove especially when validating the method. Thus, given the large sample size, even some debatable outliers were removed in order to be absolutely certain that these did not bias the calibration model or the evaluation of the prediction results. In a more realistic setting, the decision on which samples to remove could be different depending on purpose, but in this paper, the main issue is to show that the formulas work for data of good quality. Univariate regression statistics do not work well when influential samples are present, and the same holds in multivariate regression. The issue here is not finding the limit for when the formulas work but to show that they do provide meaningful results for absolutely meaningful data. Hence, we want to remove (a little too many) objects to make sure that the results are not due to an unfortunate choice of samples.

The choice of the calibration set size was based on having enough samples to adequately span the space of interferents and analyte. The remainder of the samples was then assigned as the test set. A relatively large test set is required to verify the properties of the distribution of studentised prediction residuals. Only tryptophan is discussed in the following as an example. Its concentration ranged from 0 to 7.443 μmol/l, but the concentrations were

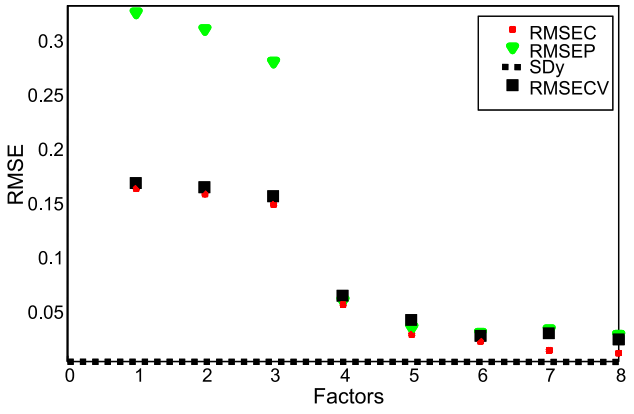

Fig. 4. Calibration errors as a function of the number of components used.

scaled such that the maximum value was 1 (for convenience of plotting). The design was purposely chosen so that some of the prediction samples were slightly outside the region of the calibration samples. This was done in order to test the approach in a demanding (extrapolation) situation (Fig. 3).

The adequacy of Eq. (4) depends directly (and almost solely) on having an accurate estimate of the calibration error. Recall that Eq. (4) is obtained by propagating random errors through the model; any systematic error in the model itself is not accounted for. The presence of significant model error shows up in, for example, a proper test set or cross-validated mean square error. If the model error is insignificant, then it follows that MSEC and MSECV should be of similar size. In case there is a significant model error, larger deviations are expected. Thus, by monitoring the gap between MSEC and MSECV, an indication of the validity of the formula can be obtained.

In this investigation, the number of components was chosen manually based on a comparison of the cross-validated and fitted predictions. In Fig. 4, the calibration (RMSEC) and cross-validation (RMSECV) results are shown for the current data problem. The prediction results (RMSEP) are also shown for convenience as well as the concentration reference uncertainty (SDy) which is situated

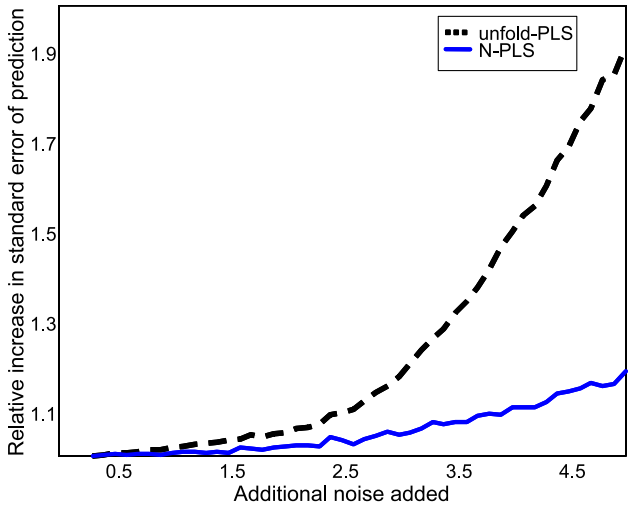

Fig. 6. Prediction standard errors for test set predictions using N-PLS and unfold PLS with different levels of noise.

almost on top of the 0 axis. A six-component model was chosen due to the minimum value of RMSECV and also due to the small gap between RMSEC and RMSECV at this number of components.

The predictions obtained with the six-component model are shown in Fig. 5.

To verify that the sample-specific standard errors of predictions can be trusted it is necessary to test that the local linearization of the error in the regression vector estimate is a valid approximation. By noise-addition as described in Theory, predictions of the test set samples are obtained for different levels of random noise added. The noise level was determined from the residuals of a six-component PCA model of the spectral data. With the added noise, six-component PLS models were determined and used to predict the test set samples.

As can be seen in Fig. 6, the standard error increases linearly with the noise level at least up to twice the intrinsic amount of noise. After this point, the added noise leads to a non-linear relation between amount of added noise and

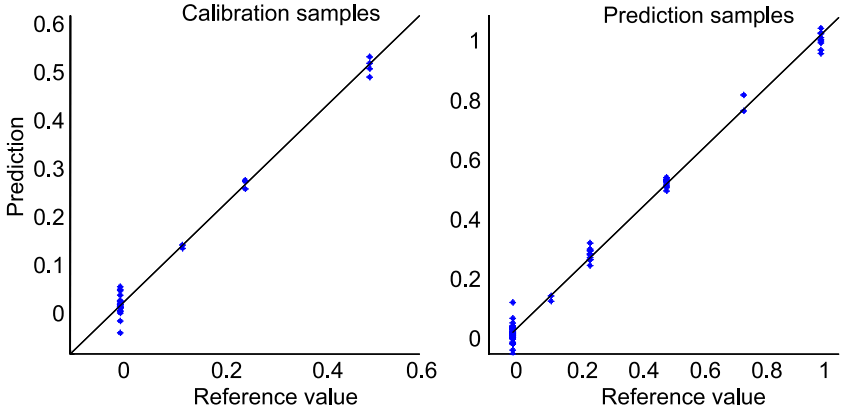

Fig. 5. Predictions for calibration (left) and test set (right). Target line superimposed.

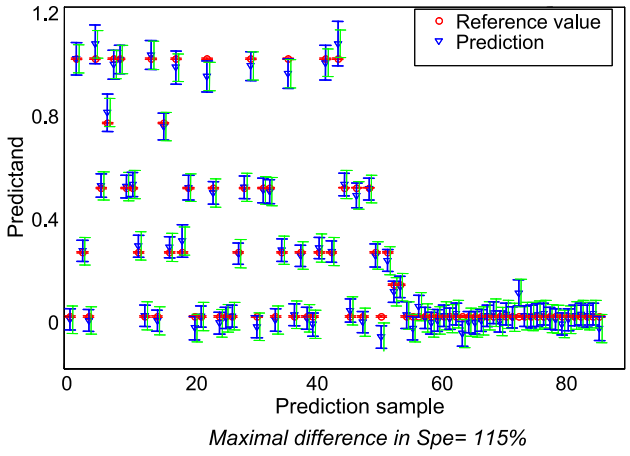

Fig. 7. Test set predictions with uncertainty. For each prediction, two 'confidence limits' are given (almost superimposed). The left-most is calculated from Eq. (4), while the right-most is twice the RMSEP. The uncertainty in the reference value is also shown but is of insignificant magnitude.

prediction standard error. Because of the linearity for low levels of added noise, the local linearization and the equations for sample-specific standard errors can be assumed to be valid from this point of view.

Fig. 6 also shows that the predictions from unfold PLS are more sensitive to the added noise than is the case for N-PLS. This is in agreement with expectations for low-noise low-rank trilinear data where the excess free parameters in unfold PLS are not needed for describing the systematic part of the data and hence lead to more noisy regression coefficient estimates.

4.2. Validation of proposed sample-specific uncertainty estimate

Using Eq. (4), the predictions in Fig. 5 (right) can be assigned individual uncertainty estimates. In Fig. 7, this uncertainty is reported as twice the estimated standard error. Also shown is the prediction interval that is calculated as twice the RMSEP. In that case, the same interval is obtained for all samples. For low concentration samples this typically (though not automatically) leads to too large intervals while the converse holds for high concentrations.

For the intervals in Fig. 7 the sample specific intervals vary by 115% as opposed to the constant size obtained from RMSEP. It is possible to assess how the sample-specific

Table 4
Percentage coverages

Level	.70	.80	.90	.95	
Theoretical	60	69	77	82	All samples
Sample-specific	62	70	76	82	
RMSEP	67	72	79	80	
Theoretical	34	39	44	47	Zero-
Sample-specific	36	40	41	46	concentration
RMSEP	40	41	46	46	samples

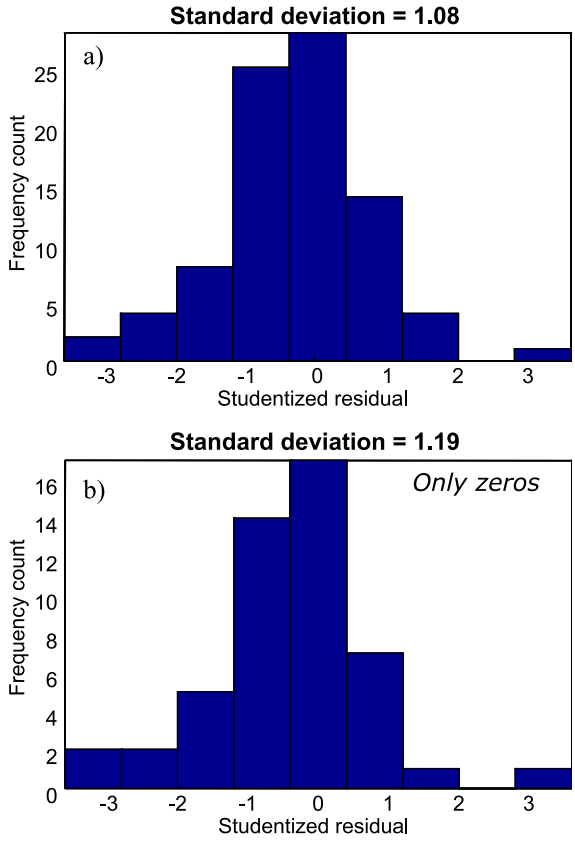

Fig. 8. Studentized prediction residuals calculated using Eq. (5) for (a) entire test set and (b) zero-concentration samples.

error coverage compares with an overall RMSEP based coverage. The set level RMSEP assumes identical prediction error but as this is not a valid assumption (compare with the univariate case), a systematic bias is expected from the RMSEP based coverage. Indeed, the RMSEP-based coverage of all samples is too high for low levels and too low for high levels (Table 4). The coverage based on sample-specific prediction errors is in better agreement with the theoretically expected results.

In Fig. 8, the studentized residuals are shown for the prediction residuals in all and for the zero-concentration sample residuals. The theoretically expected standard deviation is $\sqrt{f/(f-2)}$ where $f=35-7$ which equals 1.04. Although not perfect, the empirically observed values are close. This even holds for the zero-concentration samples; a result that is important e.g. for determining limit-of-detection. Clearly a similar approach based on RMSEP would hold little promise (Table 4).

5. Conclusions

In this paper, the approach for estimation of sample-specific prediction errors developed in Part 1 has been has been put to the test. The example shows that when the assumptions are carefully assessed to be valid, the sample-

specific errors provide a more adequate and detailed view on the prediction errors than e.g. obtained from traditional RMSEP based overall errors. The approach is based on a local linearization, but in this paper, the direct connection to earlier approaches to sample-specific prediction errors has also been highlighted. The obtained results are stimulating and point to several useful developments, e.g. for limit-of-detection estimation.

Acknowledgements

The authors gratefully acknowledge support from the Centre for Advanced Food Studies and the Danish Technical Research Council (project 1179) for financial support.

References

[1] H. Martens, T. Næs, Multivariate Calibration, Wiley, Chichester, 1989.
[2] N.M. Faber, X.H. Song, P.K. Hopke, Sample-specific standard error of prediction for partial least squares regression, Trac-Trends in Analytical Chemistry 22 (2003) 330–334.
[3] N.M. Faber, R. Bro, Standard error of prediction for multiway PLS 1. Background and a simulation study, Chemometrics and Intelligent Laboratory Systems 61 (2002) 133–149.
[4] J.A. Fernandez Pierna, L. Jin, F. Wahl, N.M. Faber, D.L. Massart, Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error, Chemometrics and Intelligent Laboratory Systems 65 (2003) 281–291.
[5] R. DiFoggio, Examination of some misconceptions about near-infrared analysis, Applied Spectroscopy 49 (1995) 67–75.
[6] D.B. Coates, Is near infrared spectroscopy only as good as the laboratory reference values? An empirical approach, Spectroscopy Europe 14 (2002) 24–26.
[7] N.M. Faber, B.R. Kowalski, Improved prediction error estimates for multivariate calibration by correcting for the measurement error in the reference values, Applied Spectroscopy 51 (1997) 660–665.
[8] L.K. Sørensen, True accuracy of new infrared spectroscopy and its dependence on precision of reference data, Journal of Near-Infrared Spectroscopy 10 (2002) 15–25.
[9] R. Wolthuis, M. van Aken, K. Fountas, J.S. Robinson, H.A. Bruining, G.J. Puppels, Determination of water concentration in brain tissue by Raman spectroscopy, Analytical Chemistry 73 (2001) 3915–3920.
[10] The American Society for Testing and Materials (ASTM), Practice E1655-00, ASTM Annual Book of Standards, vol. 03.06. ASTM, West Conshohocken, Pennsylvania, USA, 2001, pp. 573–600.
[11] A. Höskuldsson, PLS regression methods, Journal of Chemometrics 2 (1988) 211–228.
[12] PerkinElmer, Quant+ software, 2002.
[13] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Kluwer Academic Publishing, New York, 1999.