

Internal and External Validation Strategies for the Evaluation of Long-Term Effects in NIR Calibration Models

Valeria Sileoni,^{*,†} Frans van den Berg,[‡] Ombretta Marconi,[†] Giuseppe Perretti,[§] and Paolo Fantozzi[§]

[§]Italian Brewing Research Centre (CERB) and [†]Department of Economic and Food Science, University of Perugia, Via San Costanzo, 06126 Perugia, Italy

[‡]Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ABSTRACT: Some of the practical aspects of long-term calibration-set building are presented in this study. A calibration model able to predict the Kolbach index for brewing malt is defined, and four different validations and resampling schemes were applied to determine its real predictive power. The results obtained demonstrated that one single performance criterion might be not sufficient and can lead to over- or underestimation of the model quality. Comparing a simple leave-one-sample-out cross-validation (CV) with two more challenging CVs with leave-*N*-samples-out, where the resamplings were repeated 200 times, it is demonstrated that the error of prediction value has an uncertainty, and these values change according to the type and the number of validation samples. Then, two kinds of test-set validations were applied, using data blocks based on the sample collection's year, demonstrating that it is necessary to consider long-term effects on NIR calibrations and to be conservative in the number of factors selected. The conclusion is that one should be modest in reporting the prediction error because it changes according to the type of validation used to estimate it and it is necessary to consider the long-term effects.

KEYWORDS: near infrared spectroscopy, malt, brewing, Kolbach index

INTRODUCTION

Near infrared spectroscopy (NIR) is a nondestructive and rapid technique increasingly applied for food quality evaluation. It is a type of vibrational spectroscopy that employs photon energy in the range 2.65×10^{-19} to 7.96×10^{-20} J, which corresponds to the wavelength range 750 to 2,500 nm (wavenumbers 13,300 to 4,000 cm^{-1}).¹ NIR measures overtones and combinations of the molecular vibrational modes, principally those involving hydrogen atoms.² It presents many advantages over the other instrumental techniques for food and beverage analysis. For example, the low absorbance of bands is compatible with using moderately concentrated samples in combination with longer path lengths as compared to other spectroscopic techniques such as fundamental infrared (IR). These path lengths enable spectra to be measured by transmission through intact materials. Moreover, NIR spectra of intact, opaque, biological samples can be obtained by diffuse reflection so that no special cell need be used. This allows rapid, low-cost and nondestructive analysis because any form of sample preparation can be avoided or replaced by very simple procedures like grinding.³ Furthermore, NIR optical materials and the low absorbance of water, compared to e.g. IR, make the region eminently suitable for the analysis of samples with a high content of water such as foods and beverages.^{4,5} In the brewery industries, NIR spectroscopy is intensely applied to the quality control of raw materials^{6,7} and intermediate and end products.^{8,9}

However, despite the many instrumental and spectroscopic advantages, there are some issues surrounding the use of NIR in process monitoring and control in food production. NIR is an indirect semiquantitative method that is extremely reproducible under controlled conditions for getting a physiochemical fingerprint of the identity of the biological individual: the specific seed

sample.¹⁰ However, because of the high reproducibility and the low selectivity of this method, we need advanced data-handling (chemometrics) to make methods work. A direct consequence is that we need a relatively large number of training samples—both spectra and the desired reference sample—to build a regression model, and this training set should not only vary in the quantity of interest but should also “span” contributing factors such as genetics, growth conditions, seasonal variation, harvesting, storage and production technology. This especially holds for natural products such as food raw materials in open biological processes where the composition cannot be exactly predetermined. This is in contrast to humanly controlled processes where it is possible to generate synthetically composite materials based on recipes. This is why multivariate models in food processes can take years to build and/or improve up to a desirable level. The lack of selectivity in the multivariate NIR signal and the time span for model building can be problematic when other (unwanted) sources of variation are present in the data such as (small) changes in the instrumental response over time. It leads to two strict requirements in proper data analysis: correct preprocessing and model validation. The first subject is well studied and entails both hardware aspects (e.g., computing NIR absorbance values against a stable, recently collected reference material) and data treatment (e.g., removing baseline shifts by signal derivation);¹¹ the second aspect (the biological basis of data in choosing mathematical solutions) does not always get the attention it deserves and is the main subject of this paper.

Received: July 5, 2010

Accepted: January 16, 2011

Revised: January 14, 2011

Published: February 11, 2011

The aim of this work was to define a calibration model able to give a sufficiently accurate prediction of the Kolbach index (N_K) parameter by means of NIR on the flour of malt samples destined for brewing. The N_K parameter is a good indicator for the modification of the malt. It is defined as the ratio between the soluble and total nitrogen:

$$N_K = \frac{N_S \cdot 100}{N} \quad (1)$$

N_S is soluble nitrogen and N is total nitrogen, both expressed as percentage of dry matter in the flour.¹² NIR was chosen as a versatile overview of the whole physiochemical composition and as a potential substitute of the labor intensive reference method due to its ease of measurement and sample introduction, rapid response time and nondestructive nature. The task in this research was to investigate different validation strategies in the development of calibration models for the prediction of the parameter of interest, N_K , in order to define the real predictive power of the calibration and to check the long time stability of our NIR calibration models using process samples collected over different years. We emphasized model performance/validity using different figures of merit, based on different resampling schemes in order to show that one single performance criterion might be insufficient and can sometimes lead to over- or under-estimation of the model quality. The final goal was therefore to determine a realistic error of prediction, through the comparison between the results of the different validations.

MATERIALS AND METHODS

Samples. Malt samples were supplied from different industrial malt-houses and mills, and are representative of the ones available on the Italian market. 316 samples of pale malt were collected during 2006 (118), 2007 (129), 2008 (37) and 2009 (32).

Sample Preparation and Spectral Acquisition. Malt grain samples (~1 kg) were homogenized by means of a sample divider (VLB, Berlin, Germany) and finely ground by means of a DLFU type disk mill set at a distance between the disks of 0.2 mm (Bühler, Uzwil, Switzerland). The flours were used to record the spectra (Vector 22/N FT-NIR spectrometer system, equipped with a tungsten source, Rock-solid interferometer, fiber-optic module equipped with Ge-diode detector and an integrating sphere module equipped with PbS detector for spectra acquisition in diffuse reflectance mode, Bruker Optics, Milan, Italy) and to carry out the reference analyses. All $\log(1/R)$ spectra were recorded on a quartz-bottomed cup (4 cm inner diameter) placed on the integrating sphere optics and, to compensate for the lack of homogeneity, the sample was spinning during the measurement (10 rpm). Absorption spectra were collected at room temperature against a gold-coated background by means of the OPUS software (version 5.5 or 6.5, Bruker Optics) in the spectral range of 11,500–4,000 cm^{-1} (900–2500 nm) with a resolution of 8 cm^{-1} using 64 scans for averaging (the same number of scans was used for the background). In order to correct the long-term drifts affecting the reflectance/absorbance spectrum, due to changes in the water and CO_2 content in the optical path of the instrument, one reference background was collected for each sample, to calculate the sample absorbance spectrum, by using the ratio between the sample and reference signals.

NIR Instrument Validation. During the four years of this study the NIR spectrometer was subject to different checks to test the stability. On day-to-day basis a "Performance Qualification test protocol" (PQ; version 5.5 or 6.5, Bruker Optics) has been applied to ensure that the instrument was working properly. This test compares the measured data to a set of reference data which have been recorded after installation, major repair or exchange of optical components. The single tests

performed during a PQ are as follows: Deviation from 100% Line; Interferogram Peak Amplitude, Energy Distribution (Single-Channel Spectrum), X-Axis Frequency Calibration Test (Wavenumber accuracy) and Y-Axis Reproducibility Test (Photometric Accuracy). Through these tests, it is possible to detect changes in the source power. On the basis of the PQ results it was determined necessary to change the lamp every year. In this way, it was possible to avoid fluctuation in the spectra due to the natural change in the light source intensity. A new reference spectrum was measured after changing to establish a new PQ.

Reference Analyses. Standard methods from the Analytica European Brewery Convention (A-EBC) were used as reference analyses (Foss Tecator Disgestor, Foss, Hillerød, Denmark; 2200 Kjeltac Auto Distillation Unit, Foss, Hillerød, Denmark; Laborota 4600 ECO Heidelberg, Schwabach, Germany). Total nitrogen for all malts (Kjeldhal; A-EBC 4.3.1) was determined with reproducibility (R_{95}) and the repeatability (r_{95}) of the method (probability of 95%) at 0.13% m/m and 0.05% m/m respectively. The extended uncertainty, calculated from the R_{95} value following the UNICHIM guidelines for the validation of analytical chemical methods,¹⁵ was 0.1%. The Standard Error of the Laboratory (SEL), calculated from the r_{95} value following the ASTM practice E 1655-01 for quantitative NIR analysis,¹⁶ was 0.0172%.¹³ Soluble nitrogen in all malts (Kjeldhal; A-EBC 4.9.1) was determined with R_{95} and the r_{95} of the method at 0.09% m/m and the 0.04% m/m, respectively (extended uncertainty 0.07% and SEL 0.0138%).¹⁴ The R_{95} and the r_{95} for the Kolbach index determination were not available in the reference method, so the authors decided to consider the data resulting from a collaborative trial of 30 laboratories carried out by the EBC Analysis Committee in 2008 on the 17th EBC Standard Malt, in which our laboratory participated. According to these data, it was possible to estimate an extended uncertainty of 2.40 from a R_{95} value of 3.4. Moreover, it was possible to estimate a SEL of 0.41 from the r_{95} value of 1.2. Our reference values, chemically determined, ranged from 32.55 to 49.70, with a mean of 41.13. The determination range of the method is from 30 to 50, with a mean of 40.

Software. All computations involving the calibration model (spectral data pretreatments, selection of the spectral data set, construction of PLS regression model and its validation) were carried out by the MATLAB software (version 7.6) and in-house routines.

Data Analysis. A calibration model was developed using PLS1 regression.^{17,20,21} Spectra were preprocessed by extended multiple scatter correction (EMSC),^{18,19} after which the spectral range between 7501.9 and 4597.6 cm^{-1} (1330–2175 nm) was selected. Outlier samples with high error and high leverage (23 out of a total of 316) were excluded from the calibration data set. The PLS1 algorithm normally calculates several auxiliary statistics during the course of the computations. These calculations are specified by the ASTM practice E 1655-01 for quantitative NIR analysis.²⁰ One of these auxiliary statistics is the standard error of validation (SEV), more familiarly known as root mean square error prediction (RMSEP) or root mean square error of cross validation (RMSECV) in the case of cross-validation. The calculation is the same for both the parameters, and it is performed as described in eq 2 where M is number of the samples in the validation set. This parameter has been calculated for prediction using 1 to 15 PLS1 factors.

$$\text{RMSEP} = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_{i\text{-true}} - y_{i\text{-predicted}})^2} \quad (2)$$

For the chosen number of PLS1 factors, other auxiliary statistics were calculated, such as the slope of the calibration models and the bias.

In the order to check the long time stability of the NIR calibration models, four different validation test were applied:

Strategy A: The regular leave-one-sample-out cross-validation procedure, with as many one-object validation subsets as the number of samples included in the calibration set.

- Strategy B: Two more challenging cross-validation leave- N -samples-out, with $N = 73$ (25% of samples) and $N = 3$ (1% of samples), were applied in order to verify how the RMSEP values change by the number of samples used to calculate it. The selection of the samples excluded was random. Because the error will change according to chance, the two resamplings were repeated 200 times.
- Strategy C: Three test-set validations, where the validation samples were chosen according to the year of collection, were applied in the order of check if the calibration model has a stable predictive performance on samples collected in different years. Three different data blocks were defined and used for the validation: 2006, 2007, and 2008 + 2009. Two of these three data blocks were used to develop the calibration, and the third one was used as validation set. This “extrapolation” test should give a good indication of the model performance for coming years.²²
- Strategy D: Three more extreme test-set validations were applied, using the same three year based data blocks. In this case, just one data set was used as a calibration set, another block was used to find the optimal number of PLS1 factors, and the third block is used as validation set. In the first validation, samples from 2006 were used as calibration set, samples from 2007 were used for rank determination and samples from 2008 were used as validation set; this test was repeated while circulating the different year/data blocks (2006 validation, 2007 calibration, 2008 rank determination and 2006 rank determination, 2007 validation, 2008 calibration).

RESULTS AND DISCUSSION

All the barley NIR absorption spectra have been collected against a gold-coated background, which is not expected to change over time. However, trends over time could be likely connected to (small) instrumental changes. In order to check for instrumental changes over the four year period of sample collection a principal component analysis (PCA¹⁵) of the gold-coated background single beam spectra was performed (see Figure 1). The raw gold-coated background single beam spectra are shown in Figure 1a. Next, the spectral data were mean centered and a PCA was performed. From the plot in Figure 1c, where scores on the first principal component are shown as a function of collection time, we observed no obvious systematic trend. The corresponding loading vector (Figure 1b) resembles the average single beam spectrum. The second score (Figures 1b and 1d) shows a (weak) trend over time associated with a wavelength dependent change in the baseline and a noisy pattern on the NIR water-bands. This trend was unexpected, considering that during the collection time the lamp source was changed every year and several checks to test the instrumental stability have been performed, as explained in Materials and Methods. These background spectra will be used to turn flour measurements into absorbance scales, and as such it is not possible to use the observed trend directly.

To check the sample spectra a PCA was performed (see Figure 2). The raw spectral data are shown in Figure 2a; they have been mean centered before computing the PCA. Judging from the scores plot in Figure 2c the spectra of the samples collected during the year 2009 seem to differ from the others. This difference is confirmed by plotting the scores of the second principal component against the time of acquisition (Figure 2d),

where the last cluster appears to have a larger variation. This effect was unexpected, because in order to correct the long-term drifts affecting the reflectance/absorbance spectra, one reference background was collected for each sample, as explained in Materials and Methods. It could be due to a real difference of these samples compared to the others, but it is likely a more systematic (possibly instrumental or laboratory related) cause. To eliminate part of the (undesired) spectral differences, extended multiple scatter correction (EMSC) was selected as preprocessing step. This pretreatment is an expansion of the multiple scatter correction (MSC). The EMSC algorithm minimizes the signal variability caused by scatter from particulates in the samples (following the basic idea of MSC) with the inclusion of the wavelength dependency.^{18,19} A PCA analysis was performed on the pretreated data (see Figure 3). From the scores plot shown in Figure 3c we can observe that the deviation of year 2009 is reduced, possibly eliminated, compared to the other years.

All the errors of prediction obtained by cross-validations or test-set validations (RMSEPs) for 1 to 15 PLS1 factors are shown in Figure 4. Figures 4a, 4b and 4c show the RMSECV values obtained through the strategies A and B. Figures 4d, 4e and 4f show the RMSEP values calculated through the strategy C and compared with the RMSECV values calculated on the different calibration sets. Figures 4g, 4h and 4i show the RMSEP values calculated through the strategy D on the validation and the rank determination sets, compared with the RMSECV values calculated on the different calibration sets. In Table 1 are shown the most important figures of merit for the evaluation of the calibration models: RMSECV, RMSEP, bias and slope of the calibration.

The first model is based on leave-one-sample-out cross-validation (strategy A), and RMSEP is calculated for up to fifteen factors (Figure 4a). The curve has a minimum at eleven factors with a prediction error of $N_K = 1.15$. This value is lower than the extended uncertainty of the reference method, which is $N_K = 2.40$, but higher than the SEL value, which is 0.41. Checking the agreement between the calibration model and the reference method through the estimation of the confidence limit of the predicted values,¹⁶ the result is that 100% of the samples fell into the limit defined by eq 3:

$$y_{\text{predicted}} - R_{95} < y_{\text{true}} < y_{\text{predicted}} + R_{95} \quad (3)$$

To give an impression of the performance, Figure 5 shows the predicted-versus-reference plot, which has a coefficient of correlation (R -coefficient) of 0.895 for 11 PLS1 factors. From this figure it is also possible to see that all samples collected in different years show no appreciable deviations. It is also observed that samples in the high range ($N_K > \sim 45$) show a systematic, negative deviation from the predicted-versus-reference trend line. These samples will have a large influence on the slope and bias values reported in Table 1.

The number of factors (eleven) seems rather high despite the size of the data set, and it is theoretically possible but very unlikely in practice that the estimated prediction error is lower than the reference method. Leave-one-sample-out cross-validation could be too optimistic, because excluding one sample gives only a low perturbing effect on the model, while test-set validation could give a more realistic estimation of the predictive ability. The main drawback is of course the data-economic aspect: the number of samples during training/calibration is reduced. One should always remember that estimates like

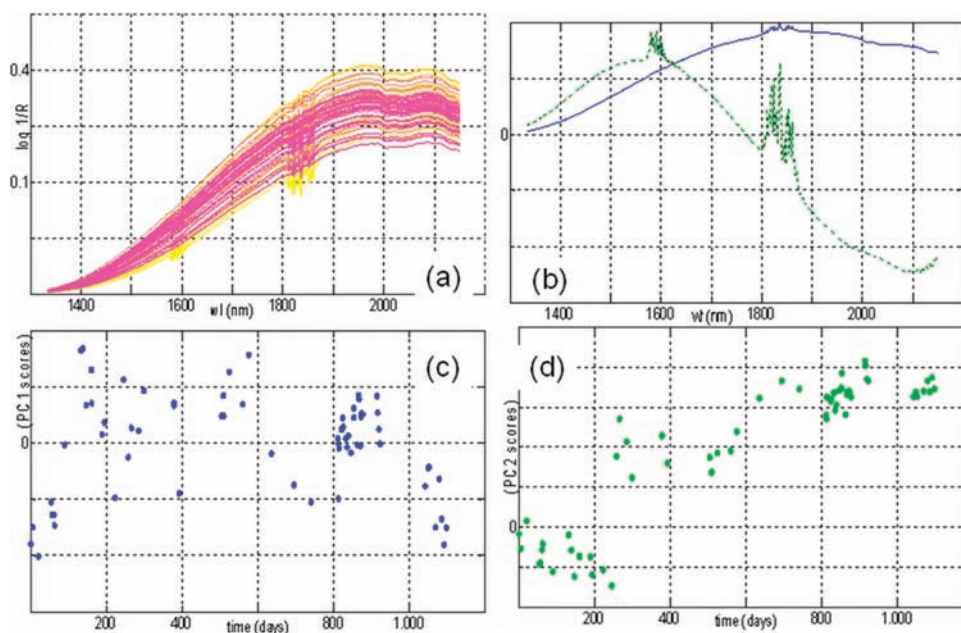


Figure 1. Principal component analysis (PCA) of selected single beam background spectra: (a) spectra, (b) PCA loadings factors 1 (—) and 2 (···), (c) scores factor 1 versus time of recording, (d) scores factor 2 versus time of recording.

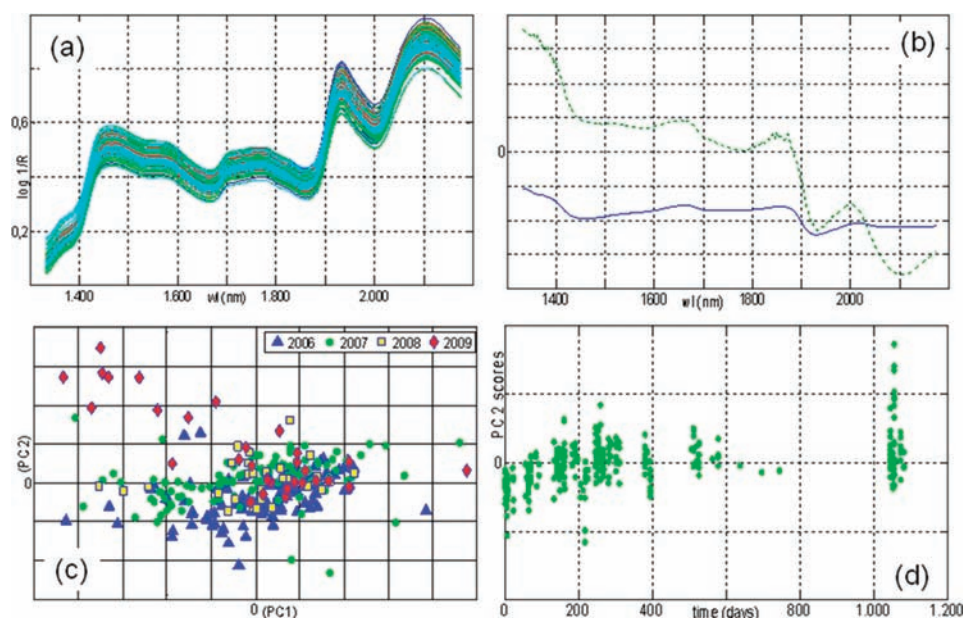


Figure 2. Principal component analysis (PCA) of raw sample spectra: (a) spectra, (b) PCA loadings factors (—) and 2 (···), (c) scores factor 1 versus scores factor 2, (d) scores factor 2 versus time of recording.

RMSEP in eq 2 are statistics, and all statistics come with uncertainty.

To get an impression of the reliability of the RMSEP values we performed a resampling CV using 200 trials selecting 25% (Figure 4b, Table 1) and 1% (Figure 4c, Table 1) at random as validation set (strategy B). The figures show the mean RMSEP value plus one standard deviation intervals from the 200 trials. Excluding 1% of the data (Figure 4c) the perturbing effect on the model is low (comparable with leave-one-sample-out CV in Figure 4a), so it is possible to obtain a better result—lower value for RMSEP—than excluding 25% of the data set (Figure 4c).

But, at the same time, we observe that the standard deviation is higher for a 1% validation set. This can be explained by some samples being very different from the others, either in true reference value or due to measurement uncertainty. This difference in samples is averaged out by increasing the validation set to e.g. 25% where the extreme samples are “hidden” among many good samples. We also observe here that the uncertainty band for leave-out-25% CV stabilizes around 5 PLS1 factors, while the RMSEP value keeps (slowly) decreasing until 10 factors.

To mimic the situation of everyday use of the NIR model in a quality control laboratory we apply the three year-blocks as test

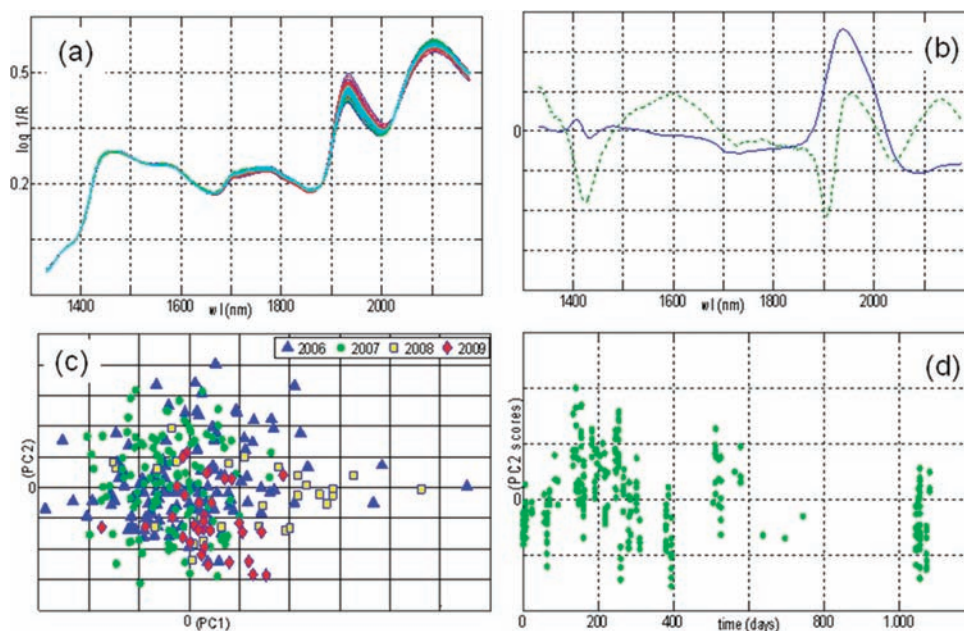


Figure 3. Principal component analysis (PCA) of EMSC preprocessed sample spectra: (a) spectra, (b) PCA loadings factors 1 (—) and 2 (···), (c) scores factor 1 versus scores factor 2, (d) scores factor 2 versus time of recording.

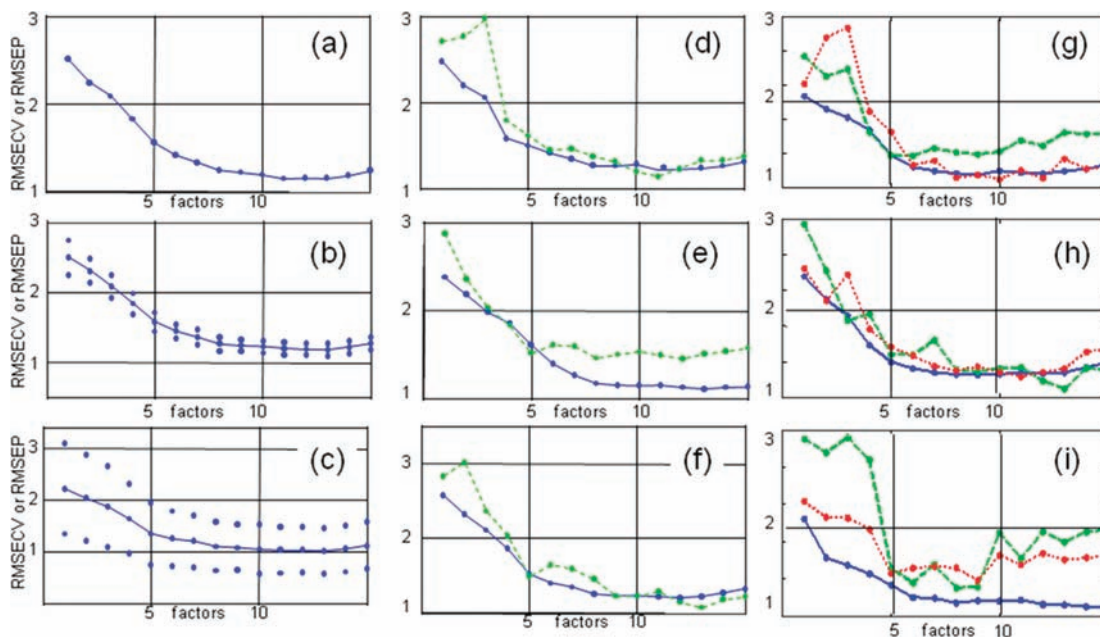


Figure 4. Prediction errors of the different validation strategies (see text and Table 1 for details).

sets of “next year” (strategy C), and the results are shown in Figure 4d–f and in Table 1. For each new calibration trail, all the samples previously excluded as outliers have been included again in the validation set and the outliers are detected by calculating the leverage and the Mahalanobis distance and estimating the confidence limits of the predicted value. The first observation is that the calibration model gives higher RMSEPs when 2007 is used as a validation set (Figure 4e). From the fifth factor on, the error of prediction calculated on the validation set is systematically higher. When 2006 is used as validation set, there is good correspondence between the calibration and validation prediction errors starting at about six factor PLS1 models (Figure 4d).

For 2008 + 2009, errors are only modestly different starting around factor five (Figure 4f). Overall the calibration model is not able to give a good prediction on the set including samples from 2007 (Figure 4e). It seems to be overfitted after the fifth component. An explanation for these worse results could be that the samples from 2007 are different from the others, even if they cannot be considered as outliers; for all the number of checked components at least the 95% of samples fall into the confidence limit of eq 3. This was double checked, but no obvious change in the instrument for this time period or exceptional/extreme samples were detected. It should thus be a more subtle connection with the samples itself, something that could possibly occur

Table 1. Results for Different Validation Strategies^a

validation strategy		samples calibration/ validation	PLS1 factors	cross-validation error RMSECV (std)	prediction error		
					RMSEP	slope	bias
A	CV leave-one-out	293/1	11	1.15	1.15	0.84	0.01
B	CV leave-25%-out (×200)	220/73	11	1.20 (0.092)	1.20 (0.092)	0.79	-0.13
	CV leave-1%-out (×200)	290/3	11	1.04 (0.451)	1.04 (0.451)	0.77	0.29
C	test set 2006 (cal “—”; val “-.-”)	179/114	11	1.21	1.13	0.83	-0.08
	test set 2007 (cal “—”; val “-.-”)	175/118	11	1.15	1.49	0.80	-0.37
D	test set 2008 + 2009 (cal “—”; val “-.-”)	232/61	5	1.60	1.52	0.61	0.11
			12**	1.13	1.46	0.82	-0.33
			11	1.22	1.29	0.78	0.45
	cal 2006 (“—”), rank 2007 (“-.-”), val 2008 + 2009 (“...”))	114/61	5	1.52	1.49	0.88	0.22
			13**	1.22	1.08	0.81	-0.18
			11	1.27	1.25	0.88	-0.74
cal 2007 (“—”), rank 2008 + 2009 (“-.-”), val 2006 (“...”))	118/114	5	1.47	1.81	0.84	1.03	
		6*	1.30	1.32	0.87	0.47	
		10**	1.22	1.12	0.88	-0.05	
	cal 2008 + 2009 (“—”), rank 2006 (“-.-”), val 2007 (“...”))	61/118	11**	1.35	1.29	0.84	-0.09
			5	1.50	1.71	0.56	0.21
			13*	1.35	1.40	0.86	0.06
			11	1.22	1.70	0.72	0.15
			5	1.42	1.58	0.56	-0.35
			8*	1.18	1.65	0.69	-0.14
			9**	1.21	1.48	0.69	-0.04

^a See Figure 4. *Minimum PCs. **Minimum RMSEP.

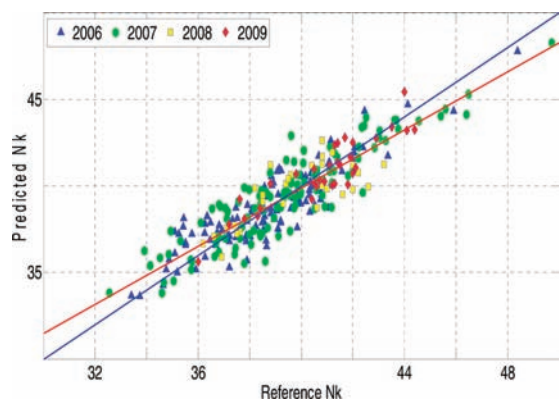


Figure 5. Leave-one-sample-out cross-validation predicted versus reference N_K values, 11 PLS1 factors with theoretical and fitted trend line.

again in future unknown samples. Factor models like PLS1 fit simultaneous variance in the predictor block (NIR spectra in this case), correlated to the reference value (optimizing covariance).^{14,17,18} PLS1 is also an eigenvalue based modeling method, so the later factors (at higher ranks), in combination with the low selectivity of NIR, make it even more difficult to identify a possible cause of the differences in behavior of different data blocks. Apparently the 2007 set has some unique features not captured by the other years once the first five factors are extracted. The first five factors appear common over all data blocks (are “robust”). The difficulty in long-term effects in NIR calibration is that the next year (2010, in our hypothesis) might look more like 2006 or 2007, and we do not know which beforehand. This might be an argument for a conservative modeling strategy, not going past five factors, sacrificing predictive performance.

The final validation challenge—strategy D where the three year-blocks are applied for calibration, model complexity selection and validation—is shown in Figure 4g–i and Table 1. Note that the calibration-set error is a pure fit error, by definition monotonically nonincreasing as a function of model complexity. In all the cases the RMSEP obtained are higher than any of the previous models, but this is expected because just one-third of the whole data set is used as calibration set to develop the model. We also observed that all three rank-determination sets suggest a PLS1 model complexity of five or six factors and the test sets RMSEPs agree well with this choice.

In conclusion, the results shown in this paper illustrate that one single performance criterion often is not sufficient to judge a near infrared calibration model in the proper way. It can lead to over- or underestimation of the model quality. Leave-one-sample-out cross validation (strategy A) is too optimistic, because excluding one sample has a low perturbing effect on the model. In our case, the confirmation of this statement is an estimated prediction error lower than the reference method. A second observation is that the root mean squared error of prediction should be considered not as a “single value” but together with its uncertainty (strategy B). In fact, using three year-blocks as test sets (strategy C) it was evident that the calibration model is not able to give an equally good prediction on the set including samples from 2007 and it seems to be overfitting after the fifth component. This is confirmed by using the same three year-based data blocks (strategy D). Of course, in all these cases the RMSEP obtained for five or six components, but also for eleven, are higher than the one obtained by leave-one-out cross-validation. The overall conclusion is that we likely need to be modest in reporting the expected prediction error.

The model developed for five components can be considered sufficiently reliable, allowing a determination of the Kolbach index parameter with a RMSEP close to the uncertainty of the reference method.

AUTHOR INFORMATION

Corresponding Author

*Tel: +39 075 973429. Fax: +39 075 973479. E-mail: vale.sileoni@tiscali.it.

REFERENCES

- (1) Osborne, B. G.; Fearn, T.; Hindle, P. H. In *Practical NIR Spectroscopy with Application in Food and Beverage Analysis*, 2nd ed.; Longman Scientific and Technical: Harlow, Essex, U.K., 1993.
- (2) De Alencar Fífueiredo, L. F.; Davrieux, F.; Flidel, G.; Rami, J. F.; Chantreau, J.; Deu, M.; Courtois, B.; Mestres, C. Development of NIRS Equations for Food Grain Quality Traits through Exploitation of a Core Collection of Cultivated Sorghum. *J. Agric. Food Chem.* **2006**, *54*, 8501–8509.
- (3) Cozzolino, D.; Kwiatkowski, M. J.; Waters, J.; Gishen, M. A feasibility study on the use of visible and short wavelengths in the near-infrared region for the non-destructive measurement of wine composition. *Anal. Bioanal. Chem.* **2007**, *387*, 2289–2295.
- (4) Woodcock, T.; Downey, G.; O'Donnell, C. P. Better Quality food and beverages: the role of near infrared spectroscopy. *J. Near Infrared Spectrosc.* **2008**, *16*, 1–29.
- (5) Huang, H.; Yu, H.; Xu, H.; Ying, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *J. Food Eng.* **2008**, *87*, 303–313.
- (6) Maertens, K.; Reyns, P.; De Baerdemaeker, J. On-line measurement of grain quality with NIR technology. *Trans. ASAE* **2004**, *47*, 1135–1140.
- (7) Marte, L.; Belloni, P.; Genorini, E.; Sileoni, V.; Perretti, G.; Montanari, L.; Marconi, O. Near infrared reflectance models for the rapid prediction of quality of brewing raw materials. *J. Agric. Food Chem.* **2009**, *57*, 326–333.
- (8) Inon, F. A.; Garrigues, S.; de la Guardia, M. Combination of mid- and near-infrared spectroscopy for the determination of the quality properties of beers. *Anal. Chim. Acta* **2006**, *571*, 167–174.
- (9) Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54* (3), 413–419.
- (10) Munck, L.; Møller Jespersen, B.; Rinnan, Å.; Fast Seefeldt, H.; Møller Engelsen, M.; Nørgaard, L.; Balling Engelsen, S. A physiochemical theory on the applicability of soft mathematical models - experimentally interpreted. *J. Chemom.* **2010**, *24*, 481–495.
- (11) Rinnan, Å.; van den Berg, F. W. J.; Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends Anal. Chem.* **2009**, *28* (10), 1201–1222.
- (12) Kunze, W. Malt production. In *Technology Brewing and Malting*, completely updated 3rd ed.; VLB: Berlin, Germany, 2004; p 177.
- (13) 4.3.2 Total Nitrogen of Malt: Kjeldahl Method. In *Analytica-EBC*; European Brewery Convention Ed.; Fachverlag Hans Carl: Nürnberg, Germany, 2006.
- (14) 4.9.1 Soluble Nitrogen of Malt: Kjeldahl Method. In *Analytica-EBC*, European Brewery Convention Ed.; Fachverlag Hans Carl: Nürnberg, Germany, 2006.
- (15) UNICHIM "Linee Guida per la validazione dei metodi analitici nei laboratori chimici" Manuale 179/1 e 179/2 Ed1995.
- (16) The American Society for Testing and Materials (ASTM) Practice E1655-00. *ASTM Annual Book of Standards*; West Conshohocken, PA, 2001; Vol. 03.06, pp 573–600.
- (17) Martens, H.; Naes, T. In *Multivariate Calibration*; John Wiley & Sons: New York, 1989.
- (18) Martens, H.; Stark, E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *J. Pharm. Biomed. Anal.* **1991**, *9*, 625–635.
- (19) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777.
- (20) Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17.
- (21) Haaland, D. M.; Thomas, E. V. Partial least-squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* **1988**, *60* (11), 1193–1202.
- (22) Sileoni, V.; Marconi, O.; Perretti, G.; Buiatti, S.; Fantozzi, P. Development of a NIR calibration model for malt extract determination and validation of its long-term stability. In *Proceedings of 9th Trends in Brewing*, Ghent (Belgium), 13–16 April 2010.