

Evaluation of basic algorithms for transferring quantitative multivariate calibrations between scanning grating and FT NIR spectrometers

Henrik Leion¹, Staffan Folestad, Mats Josefson, Anders Sparén*

Analytical Development, Pharmaceutical and Analytical R and D, AstraZeneca R and D Mölndal, SE-431 83 Mölndal, Sweden

Received 3 May 2004; received in revised form 22 September 2004; accepted 22 September 2004

Available online 23 November 2004

Abstract

A key issue in near infrared spectroscopy (NIR) is the possibility to use calibrations generated on one instrument for predictions on others. A number of methods for calibration transfer have been proposed, but which method to choose is typically not straightforward. An evaluation of a number of methods for transferring quantitative calibrations between different instruments was carried out on near infrared diffuse-reflectance data from a pharmaceutical formulation. Six instruments were included in the study, five of which were scanning grating instruments, both with and without fibre-optic probe configuration, and one of which was a Fourier-transform instrument, equipped with a fibre-optic probe. The results show that it is possible to transfer calibrations between different instruments, provided that a structured procedure is used. Simple techniques for calibration transfer, such as slope/bias correction on the predicted results, as well as standard normal variate transformation and local centring of the raw spectra, gave considerably lower prediction errors on transfer than did standardisation with a certified diffuse-reflectance standard, or direct transfer without any transfer function. Notably, including more than one instrument in the calibration also improved the prediction ability of the models on calibration transfer.

No significant differences in wavelength scale were found when a certified diffuse-reflectance wavelength standard was measured on the instruments studied. Nor did simulated wavelength scale differences below ± 0.3 nm cause any significant change in the prediction errors.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Near infrared spectroscopy; Diffuse-reflectance; Calibration transfer; Multivariate quantitative calibrations; Partial least squares regression; Solid pharmaceutical formulations

1. Introduction

Near infrared spectroscopy (NIR) has found a number of applications in the pharmaceutical industry during the last decades [1–25]. The technique is highly useful, but we believe that a prerequisite for its future success in this line of business is the ability to transfer NIR calibrations from one instrument to another. This would make NIR methods less sensitive to instrument failure and allow efficient method development on a master instrument at one site, followed by the simultaneous

use of a single calibration at different locations, using slave instruments.

During the last 15–20 years, much attention has been paid to the transfer of (multivariate) calibrations between different NIR instruments [26–29], and recently also the regulatory authorities have approached the issue [30].

There may be several reasons for the differences between spectra recorded on different instruments. However, two main types can be distinguished: *wavelength differences* originating from small differences in the wavelength scale and *response differences*, which can e.g., be in the form of an offset, a multiplicative effect, a combination of these or of a more complex nature [29]. Differences in apparent scattering may originate from minor differences in how light is coupled in and out of the analysed samples.

* Corresponding author. Tel.: +46 31 7065452; fax: +46 31 776 3768.

E-mail address: anders.sparen@astrazeneca.com (A. Sparén).

¹ Present address: Environmental Inorganic Chemistry, Chalmers University of Technology, SE-412 96 Göteborg, Sweden.

The aim of this paper was to study how a quantitative calibration for a pharmaceutical formulation developed on a single instrument (hence called the *master instrument*) could be transferred to other instruments (hence called *slave instruments*). On transfer, each of the parameters configuration (e.g., probe or sample cell) type (e.g., scanning grating or Fourier transform, FT) and brand may differ between the master and the slave instruments. In this study, type and brand coincide and therefore, only configuration and type will be discussed. Various existing methods for transferring quantitative multivariate calibrations between different NIR instruments were evaluated, aiming at:

1. Understanding how differences between NIR instruments affect the transfer of multivariate calibrations.
2. Estimating the magnitude of prediction errors introduced on the transfer of multivariate calibrations between NIR instruments, before and after applying methods for calibration transfer.
3. Outlining general strategies for transferring quantitative multivariate calibrations for solid pharmaceutical formulations between different NIR instruments.

This work is divided into two parts; the first of which is presented in this paper and evaluates a number of techniques for calibration transfer. The second part (Bergman et al., 2004; manuscript submitted to J. Pharm. Biomed. Anal.), will include a more thorough study of a selected number of methods for calibration transfer, i.e., local centring, slope/bias correction and piecewise direct standardisation (PDS).

2. Theory: algorithms for calibration transfer

Several strategies for transferring quantitative (multivariate) calibrations between different NIR instruments have

been presented. In their comprehensive reviews, de Noord [26], Bouveresse and Massart [27], Kramer [28] and Fearn [29] classified the calibration transfer methods in different ways. In this study, we focus on what is standardised and distinguish two main groups, which in their turn can be divided into several sub-groups (Fig. 1):

1. *Spectral standardisation* aims at making spectra measured on different instruments look as similar as possible. This can be achieved in various ways, such as:
 - a. Applying *spectral mapping*, i.e., transforming (mapping) spectra collected on a slave instrument to make them as similar as possible to spectra of the same samples collected on the master instrument. PDS [31,32], wavelets [33–35], finite impulse response filtering [36] (FIR) and local centring [37] are methods commonly used in this group.
 - b. Developing a *robust/invariant calibration* by selecting robust wavelength regions that do not vary much between different instruments and during the lifetime of one instrument [38], or by pre-treating spectra using e.g., standard normal variate transformation [39] (SNV), derivatives [40], multiplicative signal correction [41] (MSC) or orthogonal signal correction [42] (OSC).
 - c. *Standardising instruments*, in order to make them less sensitive to calibration transfer. In the ideal case, the differences between spectra collected on the slave and master instruments should be so small that they can be neglected. This is an important task for the instrument manufacturers.
2. *Model adaptation* aims at:
 - a. Either making the model useful for more than one instrument (*model design*) by including several instruments that should represent the instrument-to-instrument variation expected in future measurements,

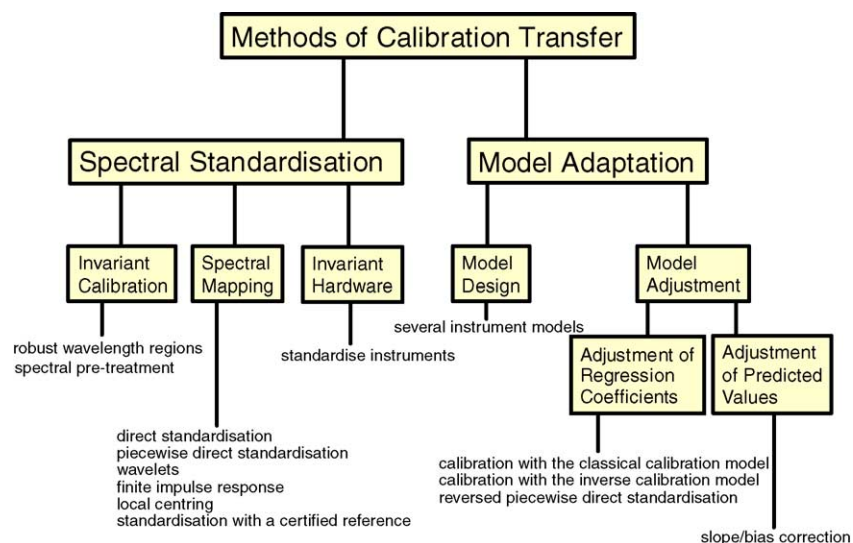


Fig. 1. Classification of methods for calibration transfer.

in order to make it robust towards instrument change and ageing.

- b. *Adjusting the model* developed on one instrument to make it work on other instruments too, which can be accomplished either by adjusting the regression coefficients of the model or by adjusting the predicted values. Calibrations using the classical or inverse calibration model [31] and reverse piecewise direct standardisation (RPDS) [43] are examples of the former approach, while slope/bias correction is an example of the latter.

This study includes methods using both the spectral standardisation and the model adaptation approaches, and the following methods in these groups were evaluated.

2.1. Spectral standardisation

Standardisation with a certified diffuse-reflectance standard, local centring, one sample as standard (spectral mapping) and SNV (invariant calibration).

2.2. Model adaptation

Including more instruments in the calibration (model design), and slope/bias correction (model adjustment).

3. Experimental

3.1. Samples

The sample set consisted of 32 coarse powder samples (AstraZeneca R&D Mölndal, Sweden), the scattering and absorptions properties of which have been characterized earlier [44]. The samples originated from 16 different batches, nine of which came from an experimental design in pilot-plant scale, while the rest of them were test batches (two in pilot-plant scale and five large-size batches). The concentration of the active ingredient varied about $\pm 25\%$ around the mean concentration in the samples. The variation originated

mainly from the samples in the experimental design. The samples were divided into a calibration set and a test set in the following way:

Calibration set

- 12 samples from the batches in the experimental design
- 6 samples from the pilot-plant test batches
- 6 samples from two large-size batches

Test set

- 8 samples from three large-size batches

None of the batches in the calibration set was used in the test set and vice versa.

Nine *certified diffuse-reflectance standards* (Labsphere, North Sutton, NH, USA), eight of which were photometric standards and one of which was a wavelength standard (WS), were also measured.

3.2. NIR measurements and instrumentation

The scanning grating instruments were of the type Foss NIRSystems 6500 (Foss NIRSystems, Silver Spring, MD, USA), equipped with either a Rapid Content Analyzer (RCA) module or an OptiProbe Fiber Optics module (probe). The FT instrument was of the type Bomem MB160 (ABB Bomem, Québec, Canada), equipped with a fibre-optic diffuse-reflectance probe.

Measurements with RCA instruments were carried out in 50-ml glass beakers, while 50-ml polyethylene flasks were used for measurements with NIR instruments equipped with fibre-optic probes. All samples were measured in triplicates on six different instruments (see Table 1), in random order. Some of the samples were also measured on a seventh instrument (F6RCA). The data from the latter were only used for an initial overview, and not in the comparison of different algorithms for calibration transfer.

Each of the certified diffuse-reflectance standards was measured once. There was no need for replicate analysis of these samples, since a feasibility study showed that ten replicate spectra of one standard gave highly repeatable results (data not shown).

Table 1
Instruments used for calibration transfer in this study

Label	Brand	Type	Configuration	Detector	Location
F1RCA	Foss	Scanning grating	RCA	Si/PbS ^b	Site A
F2RCA	Foss	Scanning grating	RCA	Si/PbS	Site A
F3Probe	Foss	Scanning grating	Probe	Si/PbS	Site A
F4Probe	Foss	Scanning grating	Probe	Si/PbS	Site B
F5Probe	Foss	Scanning grating	Probe	Si/PbS	Site B
B1Probe	Bomem	Fourier transform	Probe	InGaAs	Site C
F6RCA ^a	Foss	Scanning grating	RCA	Si/PbS	Site B

RCA, rapid content analyzer.

^a Instrument only used for the initial overview, and not in detailed comparison of techniques for calibration transfer.

^b The Si and PbS detectors are used in wavelength ranges 400–1100 nm and 1100–2500 nm, respectively. Only wavelength range 1100–2200 nm was used for the calibrations included in this study.

3.3. Reference methods

Two different in-house methods were used for reference analyses; one based on UV spectrophotometry and the other on liquid chromatography. The content of each sample was determined with either of the two methods. All reference analyses were double measurements and the mean standard deviation of these was estimated to 1.0%, according to a method proposed by Fearn [45].

3.4. Multivariate evaluation

Principal component analysis (PCA) and Partial least squares regression (PLS) modelling was carried out in Simca 8.1 (Umetrics, Umeå, Sweden). All additional calculations were made with Matlab 5.3 (MathWorks Inc., Natick MA, USA) equipped with PLS Toolbox 2.0, (Eigenvector Research, Manson, WA, USA).

The efficiency of a certain method for calibration transfer was evaluated by comparing the prediction errors obtained using the model developed on the master instrument using NIR spectra:

1. Collected on a slave instrument, applying the transfer method.
2. Recorded on the master instrument.
3. Measured with the slave instrument, without using any transfer function.

Two measures for estimating the prediction errors were used: the *root mean square error of prediction* (RMSEP), estimating the total prediction error, and the *mean bias*, estimating the systematic part of the prediction error. These measures summarise in a simple and concise way the performance of a model after calibration transfer, and the idea is similar but not identical to that suggested by Eustaquio et al. [8] for evaluating PLS models. RMSEP gives a measure of the overall prediction ability of the model, and if it increases on calibration transfer, this often shows up as an increased bias in the predictions. RMSEP and bias are defined as [46]:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i^{\text{NIR}} - y_i^{\text{ref}})^2}{n}} \quad (\text{I})$$

$$\text{bias} = \frac{\sum_{i=1}^n \hat{y}_i^{\text{NIR}} - y_i^{\text{ref}}}{n} \quad (\text{II})$$

All comparisons were made on an independent test set of samples that were not present in the calibration. RMSEP and mean bias are here presented as percent of the mean concentration of the active ingredient in the test data set, and the calculations were made for each individual replicate, i.e., triplicate measurements on one sample gave three predicted values.

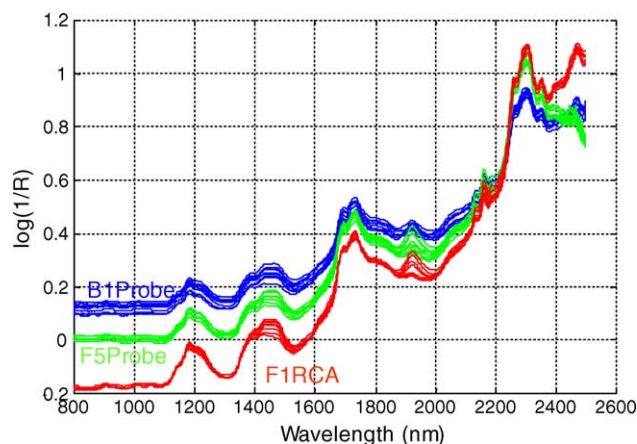


Fig. 2. NIR raw spectra obtained with three instruments of different configurations and types; a scanning grating RCA (F1RCA), a scanning grating probe (F5Probe), and an FT NIR instrument, equipped with a fibre-optic probe (B1Probe).

4. Results and discussion

4.1. Spectroscopic characterization

In Fig. 2, NIR raw spectra of the calibration set samples obtained with three NIR instruments of different configurations and types (F1RCA, F5Probe, B1Probe) are shown. It is obvious that the differences between spectra obtained with different NIR instruments were larger than the differences between spectra of samples with different content of the active ingredient. It is also clear that the differences between spectra obtained with different instruments, were not solely due to differences in offset.

It is known from the literature that differences in wavelength scale between different instruments may cause calibration transfer problems [47]. In order to check whether differences in the wavelength scale occurred for the types of instruments used in this study, a certified diffuse-reflectance wavelength standard was measured on all instruments. The results show that no significant differences in wavelength scale could be detected (Fig. 3). Further evaluation with PCA confirmed this conclusion (data not shown).

The sensitivity of the PLS models for small simulated differences in wavelength scale was also tested by shifting spectra from two of the scanning grating instruments used for prediction along the wavelength scale. The results in Fig. 3b show that differences in wavelength scale up to ± 0.3 nm had no significant effect on the total prediction error (RMSEP). Thus, as long as the scanning grating instrument meets instrument specifications, small differences in wavelength scale should not be a problem. This is in accordance with the results of Shenk and Westerhaus [47], that NIR instruments with a wavelength alignment of ± 0.25 nm is sufficient for most food and agricultural products. The wavelength accuracy of an FT instrument should be better than this, since an internal laser controls it during measurement. Specifications from

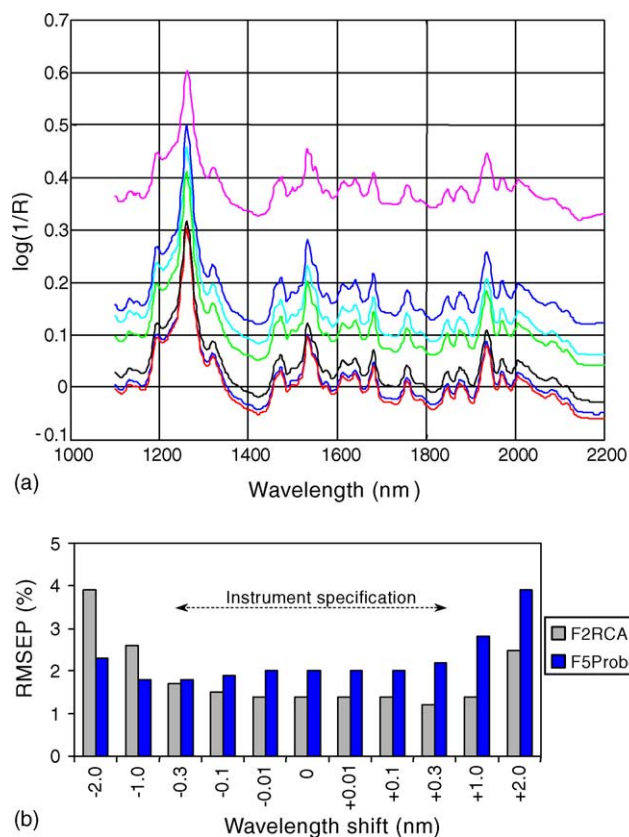


Fig. 3. The effect of small differences in wavelength scale on calibration transfer; (a) NIR spectra of a certified diffuse-reflectance wavelength standard measured on the NIR instruments included in this study; (b) the effect of simulated shifts of wavelength scale on the total prediction error (RMSEP) for PLS predictions for spectra recorded on instruments F2RCA and F5Probe, respectively.

one manufacturer reported a wavenumber reproducibility of 0.04 cm^{-1} at 7300 cm^{-1} , which corresponds to 0.015 nm at 1369.9 nm . Wavenumber repeatability (2σ) of $\pm 0.002 \text{ cm}^{-1}$ was reported [48].

4.2. Multivariate analysis

A PCA model was made on the entire data set, using a wavelength range of 1100–2200 nm. The reason for restricting the wavelength region to this range was that the FT instrument generated noisy spectra below 1100 nm, and the fact that the data from the oldest probe instruments were extremely noisy above 2200 nm. This is a commonly used wavelength range for the types of instruments studied, depending both on detector quantum efficiency and on the spectral filter of Si fibre-optics. The spectra of the FT instrument were recalculated from wavenumber to wavelength scale, in order to make them compatible with the spectra obtained using the scanning grating instruments. Spectrum resolution was adjusted to that of the scanning grating instruments.

Fig. 4 shows a PCA score scatter plot for the first two principal components, in which the instruments are grouped

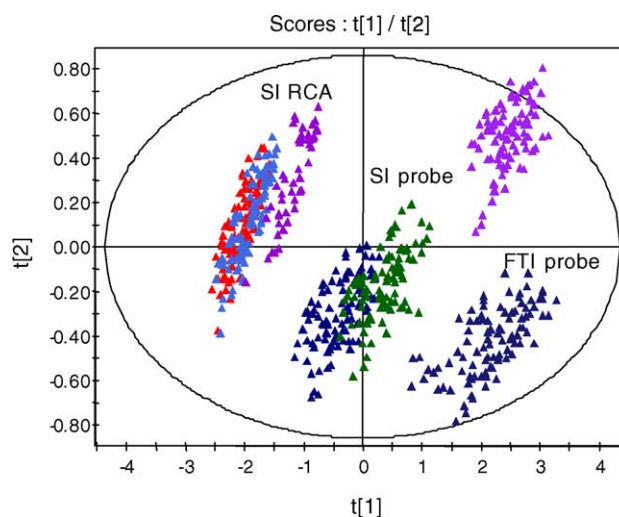


Fig. 4. A PCA score scatter plot (t_1 – t_2), on the entire data set. The spectra were mean-centred before modelling and the wavelength range 1100–2200 nm was used. Abbreviations in legends: SI, scanning grating instrument; FTI, fourier transform instrument.

into clusters that correspond to the type of instrument; the scanning grating instrument without (F1RCA, F2RCA and F6RCA) and with a fibre-optic probe (F3Probe, F4Probe, F5Probe), and the FT instrument equipped with a probe (B1Probe). Note also that the spread within each of the two groups of scanning grating instruments in the first two principal components was larger for the instruments equipped with fibre-optic probes.

The quantitative models compared below were all PLS models based on the wavelength ranges 1100–1870 and 1930–2200 nm. An initial study showed that these wavelength ranges gave the best results, in terms of few PLS components and low prediction errors. The wavelength regions excluded in the calibrations do not contain significant information of the active ingredient in the formulation.

4.3. Transfer of calibrations between instruments

4.3.1. Direct transfer

In order to test the worst-case scenario for calibration transfer, models developed on a master instrument were directly used for predicting NIR spectra measured with slave instruments, without applying any transfer function.

A calibration was developed using NIR spectra measured on a scanning grating instrument of the RCA type (instrument F1RCA, master instrument). The concentration of the active ingredient was predicted in samples from a test set measured both on the master instrument (F1RCA) and on several slave instruments of different configuration and type (F2RCA, F3Probe, F5Probe, B1Probe). The results in Fig. 5a show that the prediction errors (RMSEP, bias) obtained when using spectra recorded on a slave instrument of the same type and configuration as the one used for calibration, were in the same range as the ones obtained for the master instrument. For slave instruments of another configuration/type (probe

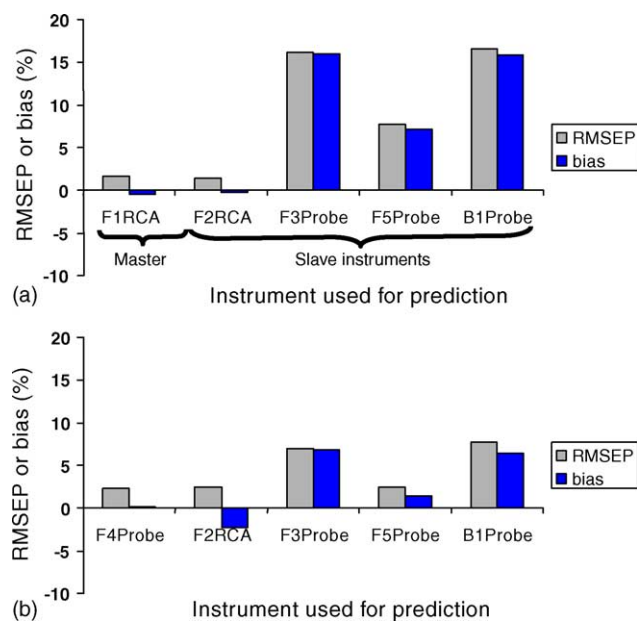


Fig. 5. Prediction errors (RMSEP, bias) obtained when predicting a test set of samples measured on the master instrument (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe). No transfer functions were used for predictions from spectra from the slave instruments. The calibration was based on (a) NIR spectra from the RCA instrument F1RCA (master instrument), and (b) NIR spectra from the scanning grating probe instrument F4Probe (master instrument).

instruments of the scanning grating and FT types, respectively), however, the prediction errors were much higher, and the bias contributed to a large portion of the total prediction error.

If the master instrument used for developing the calibration was of the scanning grating probe type (instrument F4Probe), the prediction error for the master instrument was somewhat higher (Fig. 5b) than when an RCA instrument was used as the master instrument (Fig. 5a). The prediction errors obtained when an RCA instrument was used as the slave instrument were also somewhat higher, but when the slave instrument was of the probe type, the prediction errors were considerably lower (Fig. 5) than when an RCA instrument was used as the master instrument.

A more extensive summary of results can be found in Table 2. For some calibrations, lower prediction errors were obtained for some of the slave instruments than for the master instrument. This may seem strange, but is probably just an effect of statistical uncertainty, caused by the limited number of samples used in the study. The trends are, however, trustworthy and in good agreement with the results from the coming in-depth study mentioned in the introduction, which covers a selected number of the techniques for calibration transfer, using larger calibration and test sets.

4.3.2. Spectral standardisation

4.3.2.1. Spectral mapping. It has been suggested that calibration transfer could be considerably simplified if spec-

Table 2

A summary of prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the master instrument, as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe), using various transfer functions

Master instrument	Transfer functions				
	Master	F2RCA	F3Probe	F5Probe	B1Probe
<i>No transfer function</i>					
F1RCA					
RMSEP	7.0	5.7	65.0	31.1	66.7
bias	-2.2	-0.9	64.1	28.9	63.3
F4Probe					
RMSEP	9.4	10.1	28.3	9.9	30.9
bias	0.9	-8.9	27.4	5.6	25.8
<i>Transfer function: stand certified diffuse-refl stand</i>					
F1RCA					
RMSEP	7.0	6.6	40.0	11.9	59.5
bias	-2.2	-3.3	38.5	2.5	-55.6
F4Probe					
RMSEP	9.4	11.5	34.0	13.7	28.6
bias	0.9	10.5	33.2	11.0	-23.0
<i>Transfer function: local centring</i>					
F1RCA					
RMSEP	7.0	5.8	11.1	11.7	24.4
bias	-2.2	-1.1	2.2	-0.8	-12.3
F4Probe					
RMSEP	9.4	4.8	8.1	8.2	19.2
bias	0.9	-0.3	4.0	1.2	-9.0
<i>Transfer function: SNV</i>					
F1RCA					
RMSEP	6.6	5.2	7.1	21.7	26.2
bias	-2.3	-2.2	-4.6	-20.8	-25.3
F4Probe					
RMSEP	9.2	7.4	8.1	6.9	10.6
bias	2.7	0.9	3.4	0.7	-6.0
<i>Transfer function: slope/bias correction</i>					
F1RCA					
RMSEP	7.0	5.3	9.6	7.6	9.0
bias	-2.2	-0.6	3.3	2.5	-3.7
F4Probe					
RMSEP	9.4	5.1	8.5	6.9	6.8
bias	0.9	-1.3	3.5	2.9	-1.7
<i>Transfer function: one sample as Standard</i>					
F1RCA					
RMSEP	7.0	5.7	13.3	13.7	21.1
bias	-2.2	0.1	7.7	7.2	0.2
F4Probe					
RMSEP	9.4	5.0	9.0	10.0	17.0
bias	0.9	-1.5	5.5	5.9	1.2
<i>Several instruments in calibration (no transfer functions)</i>					
F1RCA F4Probe					
RMSEP	8.7	5.0	29.7	11.2	28.1
bias	-0.4	-1.0	28.5	6.5	22.1
F1RCA F4Probe B1Probe					
RMSEP	10.7	5.8	10.3	6.5	N/A
bias	-1.1	-1.4	9.3	2.8	N/A

tra generated on a slave instrument were offset adjusted, in order to make them similar to spectra measured on a master instrument. One way of doing this would be to measure a photometric standard with a given reflectance, e.g., 80%,

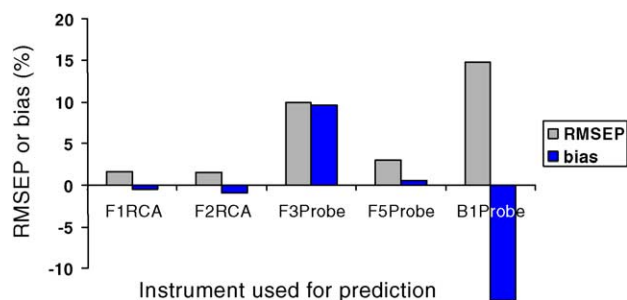


Fig. 6. Prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the master instrument (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe). Transfer function: standardization with a certified diffuse-reflectance standard. The calibration was based on NIR spectra from the RCA-instrument F1RCA (master instrument).

on both the master and the slave instruments, calculate the mean difference between the two instruments and make an offset adjustment of all spectra of samples recorded on the slave instrument, i.e., *standardisation with a certified diffuse-reflectance standard*. This implies that the only, or at least the dominating, difference between the spectra recorded on the two instruments would be a pure offset.

This idea was applied to a calibration developed on spectra from instrument F1RCA. As shown in Fig. 6, there was a slight reduction in prediction errors compared with using no transfer function, if spectra from a probe instrument were used for predicting the concentration of the active ingredient, using a calibration developed on an RCA instrument. Still, the prediction errors were large (Fig. 6). Considering Fig. 2 that clearly shows that the spectral differences between different instruments were not always pure offsets, the results are not surprising. This approach may work well for calibration transfer between very similar instruments, but then direct transfer, without the use of any transfer function, often works as well.

Local centring [37] can be a successful strategy to improve predictions if the mean spectrum of the calibration set differs from that of a new prediction set. It can also be used to account for differences in groups of spectra measured on different instruments and has been successfully used for calibration transfer [49]. The idea with local centring is that instead of mean centring a spectrum for a new sample with respect to the mean spectrum of the calibration data set measured on the master instrument, the calculation is made with respect to the mean spectrum of the group to which the new spectrum belongs, e.g., a test set measured on a slave instrument, as in this study.

The effect of using local centring for calibration transfer is shown in Fig. 7. The prediction errors were greatly reduced for all slave instruments, compared with when no transfer function was used. Still, the prediction errors obtained for predictions based on spectra recorded on the FT instrument were considerably higher than those obtained when predicting a test set on the master instrument.

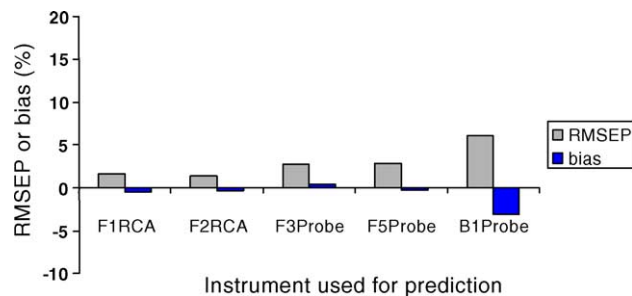


Fig. 7. Prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the master instrument (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe). Transfer function: local centring. The calibration was based on NIR spectra from the RCA-instrument F1RCA (master instrument).

A variant of standardisation with a certified diffuse-reflectance standard was also tried, in which one of the samples from the calibration carried out on the master instrument was used for standardisation of spectra recorded on the slave instrument. This technique worked well (Table 2) and can be looked upon as a variation of local centring.

4.3.2.2. Robust/invariant calibration. SNV [39] is a well-known technique for spectral pre-treatment. In NIR spectroscopy, it has been successfully used e.g., reducing light-scattering effects obtained at diffuse-reflectance measurements on solid samples. Since the technique reduces differences in offset and slope in NIR spectra, it seems likely that it should have the potential of reducing prediction errors when transferring calibration data, too.

The results obtained in this study (Fig. 8) show that SNV reduced the prediction errors of the slave instruments about as much as did local centring.

4.3.3. Model adaptation

4.3.3.1. Model design. A completely different approach to those tried so far, would be to model the differences between instruments instead of trying to minimize them before modelling. One way of doing so would be to include more than one instrument in the calibration—an idea that was tested by

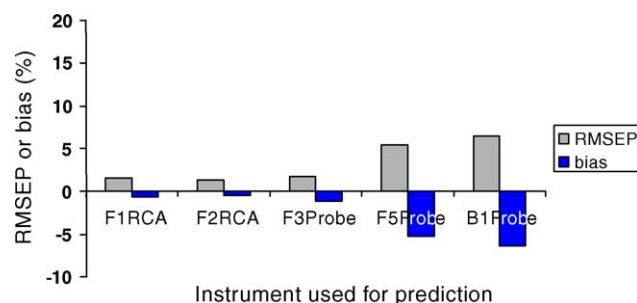


Fig. 8. Prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the master instrument (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe). Transfer function: SNV. The calibration was based on NIR spectra from the RCA-instrument F1RCA (master instrument).

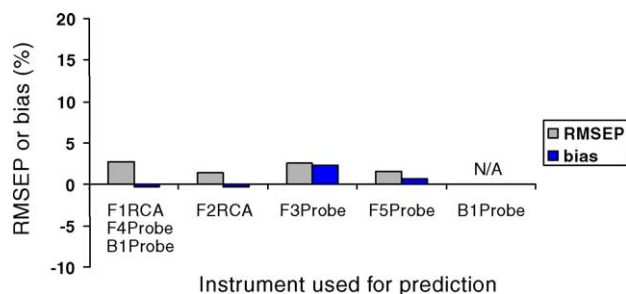


Fig. 9. Prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the three master instruments (F1RCA, F4Probe, B1Probe) included in the calibration (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe). No transfer function. Predictions for instrument B1Probe would not be relevant, since this instrument was included in the calibration, and therefore not shown in the figure.

developing PLS models based on two or three instruments. Fig. 9 shows the prediction errors obtained for three slave instruments (F2RCA, F3Probe, F5Probe) when a calibration based on spectra from three master instruments (F1RCA, F4Probe, B1Probe) of different configurations and types was used for prediction. The results show that the approach of modelling instrumental differences worked approximately as well as the best transfer functions evaluated in this study. Including more instruments in the model worked even better in combination with techniques for spectral pre-treatment, such as SNV (data not shown). Unfortunately, it was not possible to test how well the approach would work for predicting samples measured with an FT instrument, since there was only one such instrument available for the study.

4.3.3.2. Model adjustment. Slope/bias correction could be described as a simple and robust alternative to other calibration transfer techniques. The idea is to predict results, both from spectra measured on the master instrument and from those measured on slave instruments without using any transfer functions, neither on the spectra, nor on the calibration model, then applying linear regression to all pair wise values obtained at prediction with both instruments. Future multivariate predictions made from spectra recorded on slave instruments will be corrected using the linear model. The drawback of this approach is that extra samples with a large enough variation in the response variable are needed to calculate the regression equation, but since a linear relationship is assumed, a limited number of samples can normally be used. Slope/bias correction has e.g., been successfully applied to calibration transfer between a master and several slave instruments in the Danish network for NIR analysis of agricultural products [50].

The results from this study (Fig. 10) show that slope/bias correction on predicted results could be successfully applied to all slave instruments, no matter the configuration or type of NIR instrument chosen. In general, the prediction errors were lower than those obtained with any other transfer function tested. The technique can of course also be combined

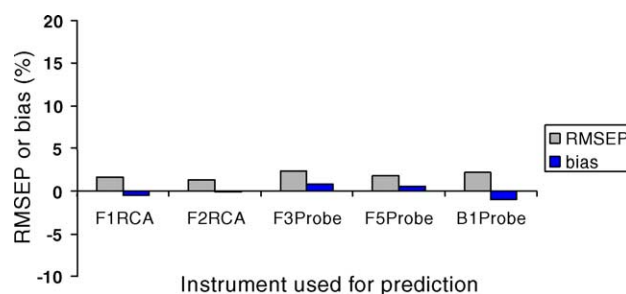


Fig. 10. Prediction errors (RMSEP, bias) obtained for predicting a test set of samples measured on the master instrument (leftmost), as well as on several slave instruments (F2RCA, F3Probe, F5Probe, B1Probe). Transfer function: slope/bias correction. The calibration was based on NIR spectra from the RCA-instrument F1RCA (master instrument).

with other techniques for calibration transfer, such as spectral standardisation.

5. Conclusions

This study shows that it was possible to transfer quantitative calibrations between different NIR instruments maintaining good prediction ability, provided that a structured procedure was used. This was true even for instruments of different configurations and types.

The prediction ability of a model after calibration transfer can be evaluated by employing two simple measures: RMSEP and mean bias. By doing so, both the total and systematic prediction errors can be estimated.

Calibration transfer between two very similar instruments, such as two scanning grating RCA instruments, may be possible without using any transfer function, while the transfer between two less similar instruments, such as those of different configuration or type, normally requires the use of some transfer function. Direct transfer of calibration data from an instrument equipped with a probe to one without a probe generated lower prediction errors than vice versa.

Including more instruments in the calibration decreased prediction errors on calibration transfer, compared with using calibrations solely developed on one instrument. It was possible to include instruments of different configurations and types in the same calibration. Even though this generated higher prediction errors for the calibration itself, it resulted in low prediction errors on calibration transfer.

Of the transfer algorithms tested in this study, slope/bias correction, SNV and local centring gave considerably smaller prediction errors on calibration transfer than did standardization with a certified diffuse-reflectance standard or direct transfer without the use of any transfer function.

No detectable differences in wavelength scale were found for any of the instruments included in the study, when measuring a certified diffuse-reflectance wavelength standard.

Simulated differences in wavelength scale up to ± 0.3 nm, which is a typical specification for scanning grating instru-

ments, gave no significant deterioration in prediction ability of the PLS models. Thus, as long as the instruments meet their specification, small differences in wavelength scale do not seem to be a problem on calibration transfer of the type of NIR methods studied here.

Acknowledgements

We thank Ola Berntsson and Lars-Göran Danielsson at the Royal Institute of Technology (Stockholm, Sweden) for kindly letting us use their FT NIR instrument, Nils-Erik Andersson, Eva-Lotta Bergman, Ingela Jedvert, Jonas Johansson, Jenny Rehn, Jonas Sjöblom, Jonas Uller and Gunnar Zackrisson for helpful discussions, Esbjörn Ljunggren for linguistic inspection and Magnus Fransson for preparing high-resolution graphic files.

References

- [1] M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH, C. de la Pezuela, *Analyst* 123 (1998) 135R–150R.
- [2] M.A. Dempster, J.A. Jones, I.R. Last, B.F. MacDonald, K.A. Prebble, *J. Pharm. Biomed. Anal.* 11 (1993) 1087–1092.
- [3] B.F. MacDonald, K.A. Prebble, *J. Pharm. Biomed. Anal.* 11 (1993) 1077–1085.
- [4] P.R. Khan, R.D. Jee, R.A. Watt, A.C. Moffat, *Pharm. Sci.* 3 (1997) 447–453.
- [5] O. Svensson, M. Josefson, F.W. Langkilde, *Appl. Spectrosc.* 51 (1997) 1826–1835.
- [6] J. Gottfries, H. Depui, M. Fransson, M. Jongeneelen, M. Josefson, F.W. Langkilde, D.T. Witte, *J. Pharm. Biomed. Anal.* 14 (1996) 1495–1503.
- [7] I. Jedvert, M. Josefson, F. Langkilde, *J. Infrared Spectrosc.* 6 (1998) 279–289.
- [8] A. Eustaquio, P. Graham, R.D. Jee, A.C. Moffat, A.D. Trafford, *Analyst* 123 (1998) 2303–2306.
- [9] P. Merckle, K.-A. Kovar, *J. Pharm. Biomed. Anal.* 17 (1998) 365–374.
- [10] A.D. Trafford, R.D. Jee, A.C. Moffat, P. Graham, *Analyst* 124 (1999) 163–167.
- [11] P. Corti, G. Ceramelli, E. Dreassi, S. Mattii, *Analyst* 124 (1999) 755–758.
- [12] M. Blanco, J. Coello, A. Eustaquio, H. Iturriaga, S. MasPOCH, *Analyst* 124 (1999) 1089–1092.
- [13] Y. Scholl, B. Draeger, *NIR News* 10 (1999) 6–8.
- [14] M.-W. Scheiwe, D. Schilling, P. Aebi, *Pharm. Ind.* 61 (1999) 179–183.
- [15] E. Dreassi, G. Ceramelli, P. Corti, P.L. Perruccio, S. Lonardi, *Analyst* 121 (1996) 219–222.
- [16] R. de Maesschalck, F. Cuesta Sánchez, D.L. Massart, P. Doherty, P. Hailey, *Appl. Spectrosc.* 52 (1998) 725–731.
- [17] O. Berntsson, L.-G. Danielsson, M.O. Johansson, S. Folestad, *Anal. Chim. Acta* 419 (2000) 45–54.
- [18] G. Radtke, K. Knop, C. Lippold, *NIR News* 10, 4 (1999) 4, 5, 12.
- [19] M. Andersson, M. Josefson, F. Langkilde, K.-G. Wahlund, *J. Pharm. Biomed. Anal.* 20 (1999) 27–37.
- [20] M. Andersson, S. Folestad, J. Gottfries, M.O. Johansson, M. Josefson, K.-G. Wahlund, *Anal. Chem.* 72 (2000) 2099–2108.
- [21] J.D. Kirsch, J.K. Drennen, *J. Pharm. Biomed. Anal.* 13 (1995) 1273–1281.
- [22] J.D. Kirsch, J.K. Drennen, *Pharm. Res.* 13 (1996) 234–237.
- [23] C. Coffey, B.E. Cooley Jr., D.S. Walker, *Anal. Chim. Acta* 395 (1999) 335–341.
- [24] A. Sparén, M. Malm, M. Josefson, S. Folestad, J. Johansson, *Appl. Spectrosc.* 56 (2002) 586–592.
- [25] J. Johansson, S. Folestad, M. Josefson, A. Sparén, C. Abrahamsson, S. Andersson-Engels, S. Svanberg, *Appl. Spectrosc.* 56 (2002) 725–731.
- [26] O.E. de Noord, *Chemom. Intell. Lab. Syst.* 25 (1994) 85–97.
- [27] E. Bouveresse, D.L. Massart, *Vib. Spectrosc.* 11 (1996) 3–15.
- [28] R. Kramer, in: *Fourth European Symposium on Near Infrared Spectroscopy: Integration into Process Control*, Biotechnological Institute Kolding, Denmark, The challenge of calibration transfer (2000) 129–135.
- [29] T. Fearn, *J. Infrared Spectrosc.* 9 (2001) 229–244.
- [30] The European Agency for the Evaluation of Medicinal Products, CPMP/QWP/160/01 (2001).
- [31] Y. Wang, D.J. Veltkamp, B.R. Kowalski, *Anal. Chem.* 63 (1991) 2750–2756.
- [32] Y. Wang, M.J. Lysaght, B.R. Kowalski, *Anal. Chem.* 64 (1992) 562–564.
- [33] G. Strang, *Am. Sci.* 82 (1994) 250–255.
- [34] B. Walczak, D.L. Massart, *TrAC Trends Anal. Chem.* 16 (1997) 451–463.
- [35] J. Trygg, S. Wold, *Chemom. Intell. Lab. Syst.* 42 (1998) 209–220.
- [36] T.B. Blank, S.T. Sum, S.D. Brown, S.L. Monfre, *Anal. Chem.* 68 (1996) 2987–2995.
- [37] A. Lorber, K. Faber, B.R. Kowalski, *J. Chemom.* 10 (1996) 215–220.
- [38] H. Swierenga, P.J. de Groot, A.P. de Weijer, M.W.J. Derksen, L.M.C. Buydens, *Chemom. Intell. Lab. Syst.* 41 (1998) 237–248.
- [39] R.J. Barnes, M.S. Dhanoa, S.J. Lister, *Appl. Spectrosc.* 43 (1989) 772–777.
- [40] A. Savitzky, M.J.E. Golay, *Anal. Chem.* 36 (1964) 1627–1639.
- [41] P. Geladi, D. MacDougall, H. Martens, *Appl. Spectrosc.* 39 (1985) 491–500.
- [42] S. Wold, H. Antti, F. Lindgren, J. Öhman, *Chemom. Intell. Lab. Syst.* 44 (1998) 175–185.
- [43] F.S.G. Lima, L.E.P. Borges, *J. Infrared Spectrosc.* 10 (2002) 269–278.
- [44] O. Berntsson, L.-G. Danielsson, S. Folestad, *Anal. Chim. Acta* 364 (1998) 243–251.
- [45] T. Fearn, *NIR News* 7 (1996) 3–3.
- [46] T. Naes, T. Isaksson, *NIR News* 2 (1991) 16–16.
- [47] J.S. Shenk, M.O. Westerhaus, *Comparison of standardization techniques*, 7th International Conference on Near Infrared Spectroscopy, Montréal, Canada, NIR Publications, Chichester, UK, 1996, pp. 112–115.
- [48] ABB Bomem Inc., MB Series FTIR Specifications and Features, Québec, Canada.
- [49] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, S. Wold, *Chemom. Intell. Lab. Syst.* 44 (1998) 229–244.
- [50] N.B. Büchmann, S. Runfors, *J. Infrared Spectrosc.* 3 (1995) 35–42.