Multi-way based calibration transfer between two Raman spectrometers

Kompany Zareh, Mohsen; van der Berg, Franciscus Winfried J

Published in:
Analyst

DOI:
10.1039/b927501k

Publication date:
2010

Document version
Early version, also known as pre-print

Multi-way based calibration transfer between two Raman spectrometers

Mohsen Kompany-Zareh*[ab] and Frans van den Berg[b]

Received 6th January 2010, Accepted 1st April 2010
First published as an Advance Article on the web 19th April 2010
DOI: 10.1039/b927501k

A standardization algorithm based on the application of Tucker3 models on the tensorized measurement signals is proposed to transfer calibration information between two Raman spectrometers. The secondary instrument in this study is a low cost and portable CCD based unit employing an efficient 532 nm green laser. The primary instrument is a high performance Fourier-transform based laboratory instrument using a low efficiency NIR laser at 1064 nm, albeit with very limited sample fluorescence interference. This work is a first investigation of calibration transfer on Raman spectral data which include different values of fluorescent background from one instrument to the other. The spectra of a small set of calibration samples are measured on both spectrometers. Using the ability of Tucker3 to estimate missing values in tensorized data, we reconstruct the spectrum of a new sample on the primary instrument based on its measured response of the secondary instrument without the need for constructing an explicit transfer model. This way spectra of a prediction sample measured on one spectrometer can be successfully transferred to another spectrometer as if it has been measured directly on the latter. Hence, the task of calibration transfer among instruments is posed as a missing data problem. A discrete wavelet transform is performed to improve the predictive ability. Performance criteria for judging the success of the calibration transfer are reported as the standard error of prediction for estimation of samples in a prediction set. By comparison, the proposed Tucker3 based standardization method shows a better performance as compared to piecewise direct standardization. The method is expected to be applicable for performing calibration transfer using data from instruments other than Raman spectrometers.

1. Introduction

Raman spectroscopy is an established method for both quantitative and qualitative determination of chemical composition and physical properties in *e.g.* food stuffs, pharmaceutical products and many other areas.[1] It is fast and nondestructive, and requires little or no sample preparation.[1–4] Development of a proper multivariate calibration model is a critical step when employing Raman spectroscopy for quantitative purposes and this normally requires considerable time and effort for preparation and measurement. Because of differences between the instrumental responses and variation of experimental conditions a practical problem in multivariate calibration occurs when an existing model (called 'primary' in our work) is applied to spectra measured under new conditions or on a different instrument (secondary).[5–9] The traditional solution to this problem consists of performing a full (re)calibration in the new situation (secondary instrument). This is obviously time and money demanding, and sometimes an experimentally burden, especially when the calibration samples are numerous, chemically unstable, hazardous, *etc.*

An alternative consists of performing a standardization, which is based on correcting the spectral difference between primary and secondary instrumental and measurement conditions. Standardization results in predicting the responses of new samples without performing recalibration, and is thus more cost-effective. Generally, the term 'standardization' encompasses several approaches such as calibration transfer, enhancement of the calibration robustness, model updating or response upgrading; this paper deals with calibration transfer. To perform the transfer of spectroscopic data, several methods for calibration transfer have been proposed in the literature. A patented algorithm proposed by Shenk and Westerhaus,[10,11] direct standardization (DS) and piecewise direct standardization (PDS) algorithms proposed by Wang *et al.*,[12–17] a two-block PLS approach suggested by Shenk *et al.*,[18] an orthogonal projection algorithm proposed by Andrew and Fearn,[19] a neural network-based approach,[20] a Fourier-based standardization method[21] and wavelet transform-based standardization techniques[22,23] are among the reported studies. In a recent study wavelet packet transform and an entropy criterion have been applied for the application of calibration transfer between two NIR instruments.[24] Zhang and Small used so-called Guided Model Re-optimization as a modification to PDS.[25] Xie and Hopke were the first to use missing values for performing calibration transfer;[26] they used the ability of the Positive Matrix Factorization algorithm in dealing with missing values.

In the present work, discrete wavelet transformation (DWT) for data compression is utilized as pre-processing method.

[a]*Department of Chemistry, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, 45137-66731, Iran. E-mail: kompanym@iasbs.ac.ir; Fax: +98-241-415-3232; Tel: +98-241-415-312*
[b]*Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark. E-mail: fb@life.ku.dk; Tel: +45-3533-3545*

Spectral data for both the primary and secondary instruments are compressed to the same level by carrying out a DWT, reconstructing the signal to the same vector length. The compressed reconstructed data form the standardization matrices for primary and secondary instruments, and a prediction matrix for secondary instrument with measured spectra. Then, applying Tucker3 on the data cube where primary and secondary matrices form the slabs, spectra measured on the secondary instrument are transferred to the primary instrument data by imputation of missing values. Calibration models are formed from the loadings in the sample direction of the Tucker3 decomposition using the known concentrations of the analytes in the sample set and loadings from the primary instrument slab. This model is then applied to predict the analytes in unknown samples using the estimated scores for the unknown samples. The full operation—decomposition of the (possibly compressed) tensor and calibration plus prediction from imputed values—will be abbreviated as T3S.

An attractive application of calibration transfer is improvement of the prediction results obtained from an ordinary instrument by using the calibration model from a high performance instrument. This is especially challenging in calibration transfer on Raman spectral data with potentially considerable differences in fluorescent backgrounds from one instrument to another. In this study, utilizing the calibration model from an FT-Raman (the high performance primary instrument) and utilizing a multi-way calibration standardization approach, prediction results from a portable CCD based instrument (the less expensive secondary instrument) were improved. Application of DWT resulted in a similar performance improvement with much less time spend in the computation step. We will compare the results with piecewise direct standardization, with and without wavelet compression. It seems that this tensorization based calibration transfer approach is applicable on other types of spectral data.

2. Theory and algorithm

2.1. Discrete wavelet transform

A transformation that uses a set of wavelet functions, localized in both spectral wavenumber axis and frequency, forms the DWT.[27,28] Stretching out the mother wavelet to fit different scales of the signal generates the wavelet basis, which is then moved to cover all parts of the signal. The wavelet transform thus provides an estimate of the local frequency content of a signal by representing the data using a family of wavelet functions that vary in scale and position. An efficient implementation of the DWT is similar to Mallet's pyramid algorithm. It performs the analysis by repeated filtering of the signal.[29] The frequency domain is cut in the middle using a pair of matched filters in each filter step. The low-frequency part is usually referred to as an approximation, the other as a detail (noise). The low-frequency components are kept in each following step, and the same matched filters are used to further subdivide the low frequency part until a given level. In this pyramid like algorithm, as described above, the details are not further analyzed and only the low frequency components are used. DWT is employed to reduce the noise and volume (vector lengths) of the data before standardization. In this work we will first augment a measurement vector to the nearest higher power of two, and then compress the signal in power-of-two steps *e.g.* the FT-Raman signal compressed to four levels would be of length 3401 → 4096 and 4096 → 2048 → 1024 → 512 → 256.

2.2. Transfer subsets

In order to transfer spectra from the secondary to the primary instrument, spectra of twenty mixture samples (solutions) of three sugars recorded from the two instruments were considered and a transfer functional description was built. A representative sample subset, which provided the most information about the response differences between the primary and the secondary instruments, was selected. To evaluate the effect of the number of standardization samples on the quality of the transformation three probes with a different number of samples in standardization subset were selected. The 20 samples were divided into a transfer set and a prediction set, according to Table 1. In the T3S method the prediction samples are analyzed one at a time where the calibration tensor is augmented by one single spectrum from the secondary instrument and one vector (tuple in a cube) with missing values. In this context, the number of samples in the prediction set is not important from a computational point of view when using T3S. However, the number of samples in the transfer part is crucial because it affects the information content and thus the quality of a model.

2.3. Piecewise direct standardization

Direct standardization creates a correction matrix to establish a mathematical relationship between the spectra from different instruments, where the response of the standardization sample on the primary instrument is directly related to the signals measured on the secondary instrument:

$$\boldsymbol{X}_i = \boldsymbol{W}_i \boldsymbol{B}_i \tag{1}$$

where $\boldsymbol{W}_i$ is the response matrix of standardization samples from the secondary instrument and $\boldsymbol{B}_i$ is the regression coefficient matrix, which can *e.g.* be calculated by means of Moore–Penrose pseudo-inverse.[5] Eqn (1) can also represent piecewise direct standardization where $\boldsymbol{X}_i$ is the estimated response vector of the primary instrument using a window i from the secondary instrument. The estimated primary responses in successive windows are then assembled to form a total estimated response matrix for the primary instrument:

$$\boldsymbol{X} = [\boldsymbol{X}_1\ \boldsymbol{X}_2 \ldots \boldsymbol{X}_\mathrm{I}] \tag{2}$$

$\boldsymbol{B}_1\ \boldsymbol{B}_2 \ldots \boldsymbol{B}_\mathrm{I}$ can be used to transfer a new spectrum collected on the secondary instrument, *e.g.* samples in a prediction set, as if they have been measured on the primary instrument. Using pseudo-inverse as a factor based method, PDS can solve ill-conditioned problems with high correlation between variables or when there are more variables than samples, both very common in spectral applications. PDS can correct intensity differences, background differences, wavelength shifts, and peak broadening, and it is therefore one of the most widely used transfer methods. To obtain an optimal transfer matrix, applying a leave-one-out

Table 1 Concentration in % (w/w) of three sugars in sample mixtures (C = calibration/transfer set, P = prediction/test set) and organizational structure for the three different probes. N is the calibration/transfer set size, $20 - N$ is the prediction set size

	Probe I ($N = 13$)	Probe I ($N = 13$)	Probe I ($N = 13$)	Sucrose	Trehalose	Glucose	
1	C	C	C	10.00	0.00	0.00	Stocks
2	C	C	C	0.00	9.58	0.00	
3	C	C	C	0.00	0.00	9.99	
4	C	C	C	0.50	0.00	0.00	Diluted stocks
5	C	C	C	0.00	0.50	0.00	
6	C	C	C	0.00	0.00	0.50	
7	P	P	P	0.51	0.49	0.50	Center-point mixture
8	C	C	P	7.98	0.98	0.00	Two component mixtures
9	P	P	P	4.96	3.87	0.00	
10	C	P	P	2.50	2.41	0.00	
11	P	P	P	0.99	1.43	0.00	
12	C	C	C	1.98	6.73	0.00	
13	P	P	P	4.49	1.46	0.00	
14	C	P	P	2.99	0.00	2.02	Three component mixtures
15	P	P	P	0.99	1.46	3.52	
16	C	C	P	0.00	1.93	5.99	
17	P	P	P	1.97	4.35	2.99	
18	C	P	P	3.99	2.89	2.99	
19	P	P	P	1.51	0.97	1.99	
20	C	C	C	2.51	1.94	0.99	

cross-validation, different window sizes ranging from 11 to 161 datapoints for the compressed data and from 51 to 1601 for uncompressed data have been evaluated using calibration transfer samples.

2.4. Tucker3 standardization

Standardization samples were measured on both the primary and secondary instruments and spectral data of prediction samples were measured only on the secondary instrument (see Fig. 1). Then, utilizing the ability of the Tucker3 algorithm to deal with missing values in the tensorized data, the prediction set spectra measured on the secondary instrument are transferred to the primary instrument conditions. Calculated sample loadings are representative of both secondary and primary instruments. The number of elements in each loading vector in the first mode ($\boldsymbol{A}$) is equal to the number of samples in standardization plus prediction (one sample at the time), as shown in Fig. 1; the number of loading vectors in all three directions is determined by the size of the core-array $\boldsymbol{G}$. The core structure/dimensionality is determined using leave-one-out cross-validation on the calibration transfer samples. A calibration model is developed on the standardization part of the sample loadings in the first mode, and sub-sequentially applied to estimate the concentration of the unknown analytes in the prediction sample. In this work we apply the Moore–Penrose pseudo-inverse to solve the multiple linear regression equation between sample loadings and reference values. This was sufficient to make the model insensitive to applied dimensionality of the core (and hence the number of vectors in loading matrix $\boldsymbol{A}$). If a factor based regression, such as PLS, had been used there would have been a need for determining the optimum number of factors.

In order to measure the model-transfer performance of various methods, the standard error of prediction (SEP), as relative expression of error, was employed:

$$\mathrm{SEP} = \Sigma_i\Sigma_j(c_{ij} - \hat{c}_{ij})^2/\Sigma c_{ij}^2 \quad (3)$$

where c_{ij} and $\hat{c}_{ij}$ are the actual and estimated concentration of analyte j ($j = 1,\ldots,3$) in the i^{th} sample in the prediction set. In case of cross-validation we will call this unit SECV, and for the calibration (or fit) error we will use the term SEC. The values will always be summed over all of the analytes unless stated otherwise. As the considered analytes in this study are three sugars, the sums are over their concentrations.

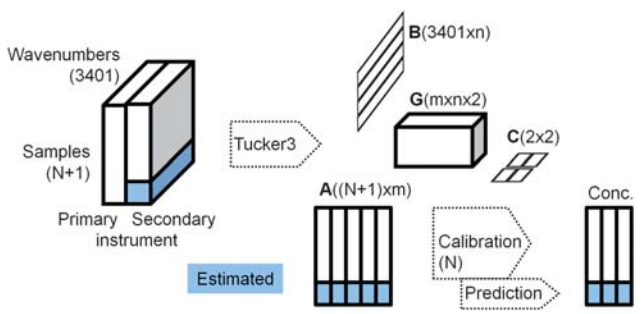

Fig. 1 T3S concept, with N standardization and 1 prediction sample at each stage.

3. Experimental

3.1. Stock and standard solutions

A simple system of sugars (sucrose [$C_{12}H_{22}O_{11}$ 342.30 g mol^{-1}] (Aldrich Chemical Co., Inc., Milwaukee, WI), trehalose [$C_{12}H_{22}O_{11}\cdot H_2O$ 378.30 g mol^{-1}] (Aldrich Chemical Co., Inc., Milwaukee, WI) and glucose [$C_6H_{12}O_6$ 180.16 g mol^{-1}] (Aldrich Chemical Co., Inc., Milwaukee, WI)) in aqueous solutions with no closure (thus the three sugars do not add up to a constant value) was used. Stock 10% (w/w) aqueous solutions of each sugar were prepared using a balance (±0.0001 g precision). Seventeen standard solutions of sucrose, trehalose and glucose were prepared according to Table 1 by weighted portions of

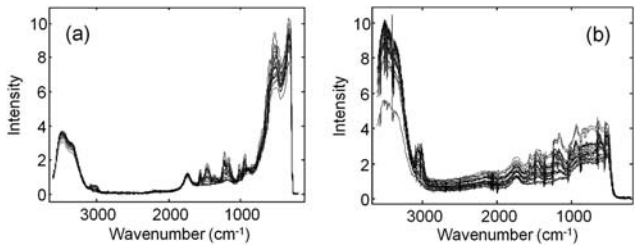

Fig. 2 Spectra of 20 samples measured on primary (a) and secondary (b) instrument.

stock solutions and water. Standards were prepared in sample tubes with a maximum weight of 7.0000 g.

3.3. Raman spectra

The 20 samples of sugars were measured on two different Raman spectrometers: the secondary instrument, a portable and inexpensive CCD based system (Ocean Optics, Dunedin, Florida, USA) and the primary instrument, a high performance FT-Raman spectrometer (Perkin Elmer System 2000, Waltham, Massachusetts, USA). The CCD based spectrometer included a probe laser with wavelength 532 nm (green), 2048 pixels, interfaced with a notebook for control and data acquisition. Maximum power of laser and liquid tip (L) of laser probe was utilized ($\sim$20 mW). Source of the FT-Raman was in the NIR region with a Nd:YAG laser line at 1064 nm. The Raman shift wavenumber range is 3600–200 cm^{-1} at 1 cm^{-1} interval. Fig. 2 shows the spectra of the samples measured on both systems.

3.4. Software

All computations were performed in MATLAB (The MathWorks, Inc., Natick, MA, USA), version 7.0. The discrete wavelet transform algorithm was written at our lab; in this work we use a db6 mother wavelet. The Tucker3 algorithm was from the PLSToolbox 4.1 (Eigenvector Research, Manson, WA). All of the other routines, such as calibration model establishment, PDS and performance evaluation, were performed with our own programs in the MATLAB 7.0 environment.

4. Results and discussion

4.1. Compression and multivariate calibration models

Spectra from the primary and secondary instruments are shown in Fig. 2; the amount of fluorescence background and noise are very different in the two systems as expected. In case of no compression the secondary instrument measurements were interpolated and adjusted to 3401 wavenumber values to match the primary instrument as close as possible (but without an algorithmic alignment/shift operation). In case of compression the median absolute deviation (MAD) statistic was applied to determine an acceptable level of compression. In the presence of a significant change in the singular values as a result of compression there would be sharp change in the MAD plot.[30,31] After converting the length of the data vectors to the closest power of two ($4096 = 2^{12}$), the datasets were evaluated at different levels where 5 and 7 were found as acceptable levels of compression for the primary and secondary instruments, respectively. Based on this 4 levels of compression (from vector length 3401 to 256) were selected for both sets to ensure that there was no significant change in the data structure.

Using spectral data from sugar mixtures as the independent variable ($\boldsymbol{X}$) and matrix of concentrations of three sugars in 20 samples as dependent variable ($\boldsymbol{y}$), without any preprocessing, PLS1 calibration models for each of the three sugars were made and leave-one-out cross-validation of calibration samples was applied to determine the optimal number of latent variables. A considerable point is that in this part the PLS regression was used to reduce the 256 variables to a limited number of latent variables in the model, however, when using limited number of loadings from Tucker3 (in the standardization stage) a simple multiple linear regression method was sufficient to form a proper regression model. Fig. 3 shows the SEP, SECV and SEC for both instruments with and without compression, using different probes. The employed regression method was PLS1 and each plot due to sum of error values for the three analytes. As it is observed from the figure, for probe III with a small number of calibration samples the SECV values are much higher than the SEP values. This is due to large differences between the samples in the small calibration set. The results show that for both instruments the optimal calibration models using any probe contained 4 latent variables (Table 2). Predictions are judged satisfactory for the primary instrument but not so for the secondary instrument, especially when the number of samples in the calibration set is reduced. The results in Table 2 also indicate that there is no significant drop in the predictive ability as a result of compression. Note that due to the lack of closure in the data and the nature of Raman measurements three separate PLS1 models (one for each sugar) with possibly different model complexities would be more sensible in practice. But, for ease of discussion, we prefer to use one model only.

4.2. Piecewise direct standardization

A piecewise direct standardization was utilized to transform the data of the secondary to the primary instrument. After the prediction sample spectra from the secondary instrument were transferred to the primary instrument, the calibration models built on the primary instrument (using 4 latent variables) were applied to the transferred spectra. The obtained prediction errors using the PDS method are given in Table 2. As expected, a lower number of samples in the calibration transfer set (going from probe I to probe III) increases the prediction errors. However, for both un-compressed and compressed datasets a considerable reduction in the prediction error was obtained using PDS. In this way, using both un-compressed and compressed data, an increase in predictive ability was observed.

4.3. Tucker3 standardization

The spectra of the prediction samples are regarded as totally missing on the primary instrument where the calibration models were built. The data from the two instruments were tensorized and a $3401 \times (N + 1) \times 2$ array was obtained (or $256 \times (N + 1) \times 2$ in the case of compression), where N is the number of transfer samples measured on both instruments. For the

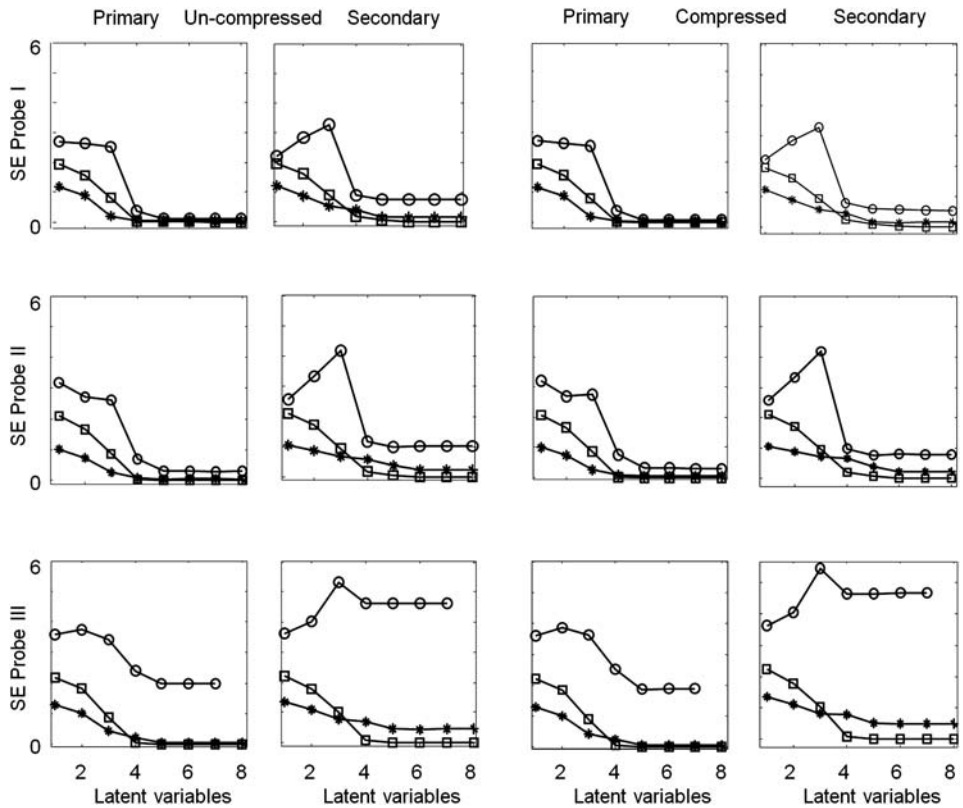

Fig. 3 Standard errors as a function of number of latent variables. SECV, SEP and SEC are assigned by circles, squares and stars, respectively.

Table 2 Standard error of prediction (SEP) values from application of PLS1 on un-compressed and compressed data from the primary and secondary instrument using the three different probes (see Table 1) and SEP values from PDS and T3S for the data from the secondary instrument. SEP values are the sum of error values for the three sugars

		Probe	Sucrose	Trehalose	Glucose	SEP[a]
Primary instrument	Un-compressed	I	0.0394	0.0068	0.0094	0.0556
		II	0.0774	0.0078	0.0082	**0.0934**
		III	0.1190	0.0595	0.0327	0.2112
	Compressed	I	0.0403	0.0073	0.0090	0.0566
		II	0.0785	0.0073	0.0064	**0.0922**
		III	0.1226	0.0663	0.0327	0.2216
Secondary instrument	Un-compressed	I	0.1136	0.0575	0.2472	0.4183
		II	0.1538	0.0700	0.3737	**0.5975**
		III	0.2366	0.3076	0.1579	0.7021
	Compressed	I	0.1233	0.0696	0.2583	0.4512
		II	0.1596	0.0891	0.4220	**0.6707**
		III	0.2390	0.3468	0.1901	0.7759
PDS	Un-compressed	I	0.0299	0.1116	0.1234	0.2650(721)[b]
		II	0.0374	0.1115	0.1339	**0.2828**(721)
		III	0.0482	0.1859	0.0677	0.3018(721)
	Compressed	I	0.0369	0.1218	0.2333	0.3920(71)
		II	0.0289	0.1300	0.0852	**0.2441**(91)
		III	0.0892	0.2320	0.0518	0.3730(101)
T3S	Un-compressed	I	0.0683	0.0588	0.0464	0.1735(7,7)[c]
		II	0.0276	0.0450	0.0481	**0.1207**(6,6)
		III	0.4450	0.6792	0.0671	1.1913(6,7)
	Compressed	I	0.0633	0.1221	0.0685	0.2539(7,6)
		II	0.0313	0.0602	0.0627	**0.1542**(6,6)
		III	0.1822	0.3276	0.0642	0.5740(7,7)

[a] 4 latent variables in all PLS predictions. [b] Applied window size in transfer step. [c] Number of factors in the mode 1 and 2 of the Tucker3 model, determined by cross-validation.

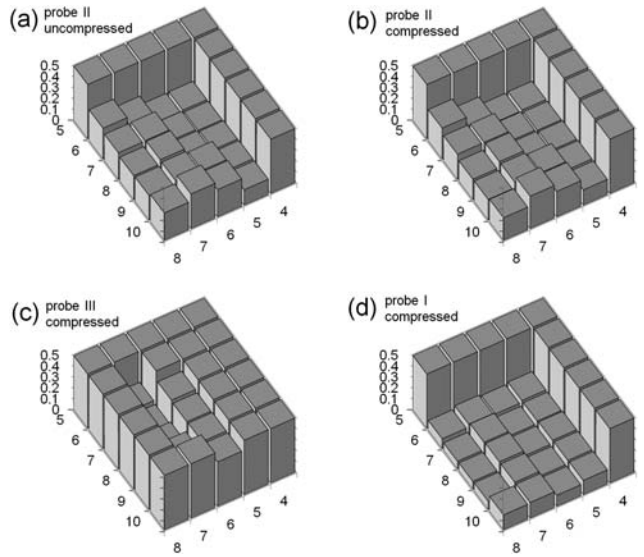

Fig. 4 SECV values (z-axis, scaled and truncated from 0 to 0.5) for the three sugars summed as a function of number of factors in mode 1 (x-axis, varying from 5 to 10) and 2 (y-axis, varying from 4 to 8) in T3S: (a) probe II no compression (min SECV = 0.1238, at (6,6) factors), (b) compressed probe II (min SECV = 0.1228 at (6,6)), (c) compression probe III (min SECV = 0.2190, at (7,7)) and (d) compression probe I (min SECV = 0.0946, at (7,6)).

N calibration transfer samples, a multivariate least squares regression was applied between the loadings in the second mode (upper part of $\boldsymbol{A}$ matrix in Fig. 1) and concentration of the compounds. Next, the estimated regression coefficients were applied to estimate the concentration of the three compounds in the unknown sample. During the Tucker3 decomposition an acceptable dimensionality of the core array in the first and second (number of loading vectors in $\boldsymbol{A}$ and $\boldsymbol{B}$ in Fig. 1) mode has to be decided. To investigate this choice further Fig. 4 shows the SECV for different selections. The number of factors in the third mode is two, which is equal to the number of slices. As can be deducted from the figure there is a relative large plateau of combinations with comparable, low SECV values. For the more challenging probe III the plateau is not very wide and the choice of the Tucker3 core is critical. It is worth noting that the time spent for running Tucker3 on the compressed data is considerably lower than the time required for un-compressed data due to the imputation step. SEP values at the optimum number of factors obtained from SECV values (using a leave-one-out cross-validation) are listed in Table 2.

To evaluate the T3S procedure Fig. 5 presents the difference between the transformed and measured spectra on both primary and secondary instruments (probe II). It can be seen from Fig. 5a that the spectral differences between primary and secondary data are considerably reduced after standardization and are both positive and negative, largely concentrated below ~900 cm^{-1}. Higher values of residuals are in the regions with less spectral similarity between secondary and primary instruments. The predicted concentrations were plotted against the reference values for some example situations (Fig. 5c and d); after the application of T3S calibration transfer good correlation was observed.

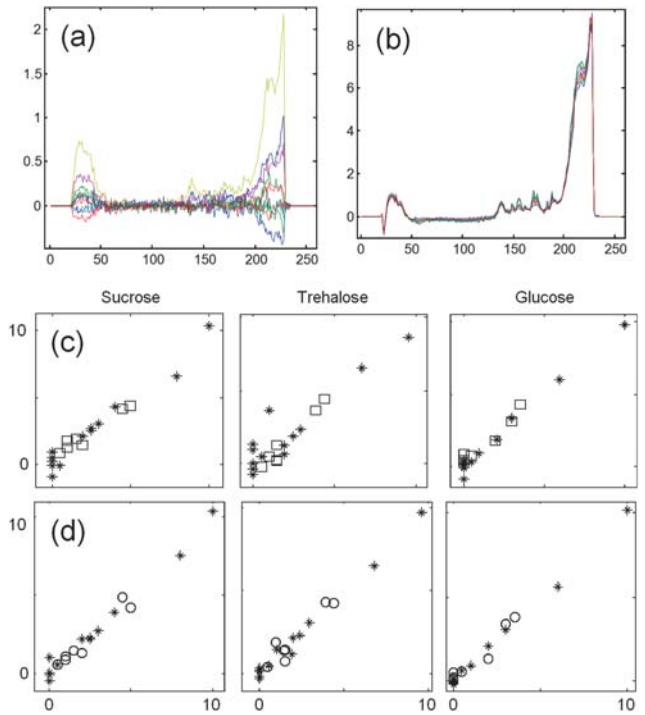

Fig. 5 Probe II compressed data with T3S estimation: (a) difference between spectra from primary instrument and estimates for secondary instrument, (b) estimated spectra for secondary instrument. Estimated *versus* reference concentration in calibration (*) and test (○) set separated by sugar for (c) un-compressed probe II data of secondary instrument (SEP = 0.5975) and (d) compressed probe II plus T3S (SEP = 0.1542).

5. Conclusion

In this investigation the problem of calibration transfer is considered as a missing data problem. All spectra of the prediction samples on the primary instrument were taken as missing values. These missing spectra were then predicted by the proposed Tucker3 based approach and calibration models built on the loading from the calibration part could be applied. Employing the proposed Tucker3 based method no explicit transfer matrix is calculated. The transferred spectra were estimated through a data matrix reconstruction by Tucker3; the combined steps are called T3S. The results demonstrated the feasibility of using T3S for instrument standardization. An improved performance of the T3S algorithm over a PDS method on two Raman spectrometers of highly different quality was illustrated. A compression of the data by discrete wavelet transform did not result into a better prediction, but reduced the time consumption for calculations by imputation considerably without loss of performance (going from ~30 seconds to less then 5 seconds on a process computer, enabling "real-time" process monitoring). When the number of times of measurements and computations are very large this reduction in processing time becomes remarkable.

Acknowledgements

This work was supported by the Institute for Advanced Studies in Basic Sciences (Zanjan, Iran) grant no. G2009IASBS123 and

The Danish Research Council for Technology and Production Sciences (project "Hard-model based tensor computations in the study of small molecules interactions with DNA").

References

1 N. Viereck, T. Salomonsen, F. W. J. van den Berg and S. Balling Engelsen, Raman Applications in Food Analysis, in *Raman Spectroscopy for Soft Matter Applications*, ed. M. S. Amer, Wiley Interscience, MA, USA, 2009, pp. 199–223.
2 S. Ch. Park, M. Kim, J. Noh, H. Chung, Y. Woo, J. Lee and M. S. Kemper, *Anal. Chim. Acta*, 2007, **593**, 46–53.
3 F. C. C. Oliveira, C. R. R. Brandão, H. F. Ramalho, L. A. F. da Costa, P. A. Z. Suarez and J. C. Rubim, *Anal. Chim. Acta*, 2007, **587**, 194–199.
4 M. F. Mrozek, D. Zhang and D. Ben-Amotz, *Carbohydr. Res.*, 2004, **339**, 141–145.
5 F. W. J. van den Berg and Å. Rinnan, Calibration Transfer Methods, in *Infrared Spectroscopy for Food Quality Analysis and Control*, ed. D.-W. Sun, Elsevier, USA, 2009, pp. 105–118.
6 H. W. Tan and S. D. Brown, *J. Chemom.*, 2001, **15**, 647–663.
7 F. Despagne, D. L. Massart, M. Jansen and H. V. Daalen, *Anal. Chim. Acta*, 2000, **406**, 233–245.
8 E. Bouveresse and D. L. Massart, *Vib. Spectrosc.*, 1996, **11**, 3–15.
9 E. Bouveresse and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1996, **32**, 201–213.
10 J. S. Shenk and M. O. Westerhaus, *U.S. Pat.*, 4 866 644, Sep. 12, 1991.
11 E. Bouveresse, D. L. Massart and P. Dardenne, *Anal. Chim. Acta*, 1994, **297**, 405–416.
12 Y. Wang, D. J. Veltkamp and B. R. Kowalski, *Anal. Chem.*, 1991, **63**, 2750–2756.
13 E. Bouveresse, D. L. Massart and P. Dardenne, *Anal. Chem.*, 1995, **67**, 1381–1389.
14 Y. Wang, M. J. Lysaught and B. R. Kowalski, *Anal. Chem.*, 1992, **64**, 562–564.
15 J. Lin, S. C. Lo and C. W. Brown, *Anal. Chim. Acta*, 1997, **349**, 263–269.
16 A. Herrero and M. C. Ortiz, *Anal. Chim. Acta*, 1997, **348**, 51–59.
17 F. Sales, M. P. Callao and F. X. Rius, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 63–73.
18 J. S. Shenk and M. O. Westerhaus, *NIR News*, 1993, **4**, 13.
19 A. Andrew and T. Fearn, *Chemom. Intell. Lab. Syst.*, 2004, **72**, 51–56.
20 F. Despagne, B. Walczak and D. L. Massart, *Appl. Spectrosc.*, 1998, **52**, 732–745.
21 C. S. Chen, C. W. Brown and S. C. Lo, *Appl. Spectrosc.*, 1997, **51**, 744–754.
22 B. Walczak, E. Bouveresse and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1997, **36**, 41–51.
23 J. Yoon, B. Lee and C. Han, *Chemom. Intell. Lab. Syst.*, 2002, **64**, 1–14.
24 Ch. Tan and M. Li, *Anal. Sci.*, 2007, **23**, 201–206.
25 L. Zhang and G. W. Small, *Anal. Chem.*, 2003, **75**, 5905–5915.
26 Y. Xie and P. K. Hopke, *Anal. Chim. Acta*, 1999, **384**, 193–205.
27 B. Walczak and D. L. Massart, *Chemom. Intell. Lab. Syst.*, 1997, **36**, 81–94.
28 B. Walczak, B. V. Bogaert and D. L. Massart, *Anal. Chem.*, 1996, **68**, 1742–1747.
29 H. W. Tan and S. D. Brown, *J. Chemom.*, 2001, **16**, 228–240.
30 E. R. Malinowski, *J. Chemom.*, 2009, **23**, 1–6.
31 M. Kompany-Zareh and F. W. J. van den Berg, *Anal. Chim. Acta*, 2010, DOI: 10.1016/j.aca.2010.04.017.

