MODELLING TEMPERATURE-INDUCED SPECTRAL VARIATIONS IN CHEMICAL PROCESS MONITORING

Zeng-Ping Chen, Julian Morris, Elaine Martin

Centre for Process Analytics and Control Technology,
School of Chemical Engineering and Advanced Materials,
University of Newcastle upon Tyne, NE1 7RU, UK

Abstract**:** With a view to ensuring the validity of multivariate calibration models in the presence of temperature variations, a new methodology, individual contribution standardization (ICS), is proposed to correct for temperature-induced spectral variations. The methodology was applied to shortwave NIR spectral data sets recorded at different temperatures. The results showed that ICS can almost eliminate temperature effects resulting in calibration models that exhibit good predictive performance. Compared with other methods, such as continuous piecewise direct standardization, ICS has the advantages of easy implementation, simple selection of the calibration model and less restrictions on the training samples. *Copyright© 2004 IFAC*

Keywords: Chemical process monitoring, Temperature-induced spectral variations, Multivariate calibration model.

1. INTRODUCTION

Spectroscopic techniques are increasingly being applied in process analysis (Blaser, et al., 1995; Lewiner, et al., 2001) as a result of their potential for on-line application and the short analysis time. Due to the lack of selectivity of the target species, in many chemical processes, multivariate methods such as Partial Least Squares (PLS) (Höskuldsson, 1988) are applied for the building of robust calibration models. Generally these models are valid only when the spectra of the future samples (test samples) are measured at the same external conditions as the calibration samples (training samples). However in industrial applications, on-line spectral process measurements are often affected by fluctuations in external variables such as temperature, thereby invalidating the multivariate calibration model if the effects of the external variations on the spectral measurements are not taken into account when developing the model.

One approach proposed for addressing external sources of variation has been to implicitly incorporate the effect of temperature (or other variables) into the calibration experimental design. Wülfert, et al. (1998) studied the effect of fluctuating temperature on the predictive ability of a multivariate calibration model for NIR spectra. They concluded that by implicitly including temperature in the model, satisfactory predictive performance was observed, although the complexity of the model was increased. As neither the temperatures of the calibration samples nor that of the future samples is included explicitly in the model as independent or dependent variables, the accuracy of the results will not be optimal.

An alternative approach to the handling of temperature influences is through the explicit inclusion of temperature into the calibration model by treating the temperature of the samples as an extra independent variable appended to the spectra or as another dependent variable, for example. Wülfert, et al., (2000a) investigated this approach. Their results showed that none of the approaches that explicitly include temperature in the calibration model results in an improvement in the performance of the calibration model when compared to the more basic idea of implicit inclusion.

It was proved that non-linearities such as temperature effects (Iwata, et al., 1997) cannot be filtered out or resolved through the application of an orthogonal basis transformation (Wülfert, et al., 2000a), such as the wavelet transformation. An alternative approach has been to use robust variable selection models to decrease the influence of temperature variations on a model's predictive performance (Swierenga, et al., 2000). It was observed that models based on robust variable selection may be better than, or similar to, implicit global models with respect to prediction errors. However, a disadvantage related to variable selection models is that there is a need for an understanding of the underlying chemical system.

Wang, et al. (1991, 1993) developed piecewise direct standardization (PDS) to correct for complex non-linear spectral variations to enable calibration transfer between measurements performed on two different instruments or under two different sets of conditions (say two different temperatures). However, a requirement is that measurements are made for a selected subset of samples for both the calibration and test temperatures, for example, thereby making the approach unsuitable for correcting temperature-induced spectral variations in process analysis, since it is not possible to measure the spectra of a selected subset of samples under all possible temperatures since temperature is a continuous variable.

Continuous piecewise direct standardization (CPDS) (Wülfert, et al., 2000b) is a generalization of PDS to continuous variables. To overcome the restriction of PDS, CPDS fits polynomials to the corresponding elements of the discrete PDS transformation matrices under different temperatures. The transformation matrix for a new temperature can then be estimated from the established polynomials. CPDS can provide results that are superior to the implicit or explicit inclusion of temperature into the calibration model.

Three parameters (the number of latent variables, the width of the window size and the degree of the polynomial) have to be estimated to build the CPDS correction model. The optimization procedures are complex and time consuming. The requirement that the spectral measurements are obtained for exactly the same subset of samples at all calibration temperatures cannot always be satisfied in process analytical applications. Moreover, the corresponding elements of the discrete PDS transformation matrices under different temperatures may not follow smooth non-linear models. The difference between the transformation matrix estimated for a new temperature by the polynomial models established and the actual one may not be trivial, hence the resulting prediction would be inaccurate.

In this paper, a new methodology is proposed, **I**ndividual **C**ontribution **S**tandardization (ICS), for the correction of temperature-induced spectral variations. The method is evaluated using a NIR data set and the results are compared with CPDS.

2. THEORY AND ALGORITHM

Suppose the rows of the spectral data matrices $\mathbf{X}(t_1)$, $\mathbf{X}(t_2)$,..., $\mathbf{X}(t_K)$ are the corresponding spectra of the training mixture samples measured at training temperature t_1, t_2, …, and t_K, respectively with known concentration matrix **Y**. The objective is to build multivariate calibration models that can predict the concentrations of the chemical components in future mixture samples from the spectral data measured at certain temperatures. If the temperature of future mixture sample, t_k, is equivalent to any of $t_1, t_2, \ldots, t_K$, a prediction can be made using the local model built between $\mathbf{X}(t_k)$ and $\mathbf{Y}$.

However temperature, as stated, is a continuous variable. It can therefore take any value. Due to the difference in temperature between the future samples and the training samples, the local model built on the training samples cannot give correct predictions for future samples if the temperature-induced spectral variations have not been appropriately corrected for.

To correct for the influence of temperature, the continuous piecewise direct standardization (CPDS) algorithm (Wülfert, et al., 2000b) adopts the following strategy. First a partial least square (PLS) calibration model between the spectral data $\mathbf{X}(t_{ref})$ and the concentration matrix $\mathbf{Y}$ is established for a selected reference temperature t_{ref} from the training temperatures. The next step is the calculation of the discrete calibration transformation matrices between the spectra measured at the remaining training temperatures and the reference temperature. These are found through piecewise direct standardization (PDS) (Wang, et al., 1991):

$$\mathbf{X}(t_{ref}) = \mathbf{X}(t_k)\mathbf{P}(t_k)\,, \quad (k = 1,2,\cdots,K;\, t_k \neq t_{ref}) \qquad (1)$$

The transformation matrix $\mathbf{P}(t_k)$ is a banded matrix that can correct for the spectral variations resulting from the temperature differences between two distinct measurement temperatures. The transformation matrix $\mathbf{P}(t_k)$ is obtained by regressing the elements, $\mathbf{x}_j(t_{ref})$ (wavelength *j*, reference temperature t_{ref}) on a window $\mathbf{x}_{j-w}(t_k)$ to $\mathbf{x}_{j+w}(t_k)$. The widow width is $(2\times w+1)$. The regression vectors $\mathbf{b}_j(t_k)$ are then calculated using PLS and the transformation matrix $\mathbf{P}(t_k)$ is constructed by placing $\mathbf{b}_j(t_k)$ in the *j*th column, that ranges from row $j-w$ to $j+w$. A polynomial is then fitted to each element along the bands of the main diagonal of the transformation matrices $\mathbf{P}(t_k)$ $(k=1,2,\cdots,K;t_k \neq t_{ref})$ against the temperature difference Δt_k ($\Delta t_k = t_k - t_{ref}$):

$$p_{i,j}(t_k) = a_{i,j} + b_{i,j}\Delta t_k + c_{i,j}\Delta t_k^2 + e_{i,j} \tag{2}$$

Once the parameters of the polynomials have been estimated, the transformation matrix $\mathbf{P}(t_{test})$ for the spectral matrix $\mathbf{X}(t_{test})$ of the test samples measured at temperature t_{test} can be calculated. The temperature effects can then be removed:

$$\mathbf{X}(t_{test} \rightarrow t_{ref}) = \mathbf{X}(t_{test})\mathbf{P}(t_{test}) \tag{3}$$

The concentrations can be predicted by using the local model at the reference temperature. To obtain satisfactory results, three parameters have to be optimized prior to building the final CPDS model, the number of latent variables for calculating the regression vector $\mathbf{b}_j(t_k)$, the width of the bands about the main diagonal (window size $(2\times w+1)$) and the degree of the polynomials to fit the elements of the transformation matrices. The best values for the three parameters are determined by leave-one-out cross validation. This approach is time consuming. Furthermore, the selection of the reference temperature is critical for accurate predictions.

In their original paper, Wülfert, et al. (2000b) proposed selecting a midpoint calibration temperature for better results. Intuitively, for test samples measured at different temperatures, the best reference temperatures should be different. The smaller the difference between the reference and test temperature, the better the results. However, building CPDS models for all calibration temperatures is challenging. The optimal values for the three parameters may not be the same when different reference temperatures are considered. Consequently considerable time has to be spent optimising these parameters for each reference temperature. Even if the optimal values for the three parameters are found, the elements on the bands of the transformation matrices may not follow smooth non-linear functions. Hence, the results of CPDS may not be reliable.

Generally, temperature has a different impact on different chemical species in mixture samples, and the absorbance of each chemical species at different wavelengths will follow simple monotonic smooth non-linear functions with respect to temperature. These facts clearly indicate that the temperature-induced spectral variations should be modelled separately for each chemical species in mixture samples. Consequently the individual contribution standardization (ICS) approach developed adopts the following strategy. The response patterns $\mathbf{s}_l(t_k)$ $(l=1,2,\cdots,m)$ of chemical species l under temperature t_k can be resolved using the following direct regression approach:

$$\mathbf{S}(t_k) = \mathbf{Y}^{+}\mathbf{X}(t_k)$$

$$\mathbf{S}(t_k) = \begin{bmatrix} \mathbf{s}_1^{'}(t_k) \\ \mathbf{s}_2^{'}(t_k) \\ \vdots \\ \mathbf{s}_m^{'}(t_k) \end{bmatrix}, \quad (k=1,2,\cdots,K) \tag{4}$$

A polynomial is then fitted for each element of $\mathbf{S}(t_k)$ $(k=1,2,\cdots,K)$ against temperature t_k.

$$s_{i,j}(t_k) = a_{i,j} + b_{i.j}t_k + c_{i,j}t_k^2 + e_{i,j} \tag{5}$$

After $a_{i,j}$, $b_{i,j}$ and $c_{i,j}$ have been estimated, the response pattern matrix $\mathbf{S}(t_{test})$ of the chemical species under test temperature t_{test} can be calculated. Given the spectral matrix $\mathbf{X}(t_{test})$ of test samples measured under temperature t_{test}, the contribution of the chemical species can be approximately estimated:

$$\hat{\mathbf{Y}} = \mathbf{X}(t_{test})\mathbf{S}^{+}(t_{test}) \tag{6}$$

As a result of the existence of interferences such as background, $\hat{\mathbf{Y}}$ is not necessarily an acceptable estimate of the concentration matrix for the test samples. However, it can be used to transform $\mathbf{X}(t_{test})$ to $\mathbf{X}(t_{test} \rightarrow t_{ref})$, as though it were measured under reference temperature, t_{ref} :

$$\mathbf{X}(t_{test} \rightarrow t_{ref}) = \hat{\mathbf{Y}}(\mathbf{S}(t_{ref}) - \mathbf{S}(t_{test})) + \mathbf{X}(t_{test}) \tag{7}$$

Here, $\mathbf{S}(t_{ref})$ is the response pattern matrix of the chemical species for reference temperature t_{ref} . $\hat{\mathbf{Y}}(\mathbf{S}(t_{ref}) - \mathbf{S}(t_{test}))$ can be viewed as the difference introduced by the temperature variation between $\mathbf{X}(t_{test})$ and $\mathbf{X}(t_{test} \rightarrow t_{ref})$.

After the spectral variation introduced by the difference in temperatures has been corrected for, calibration models established under reference temperatures can be developed to predict the concentration of all the chemical species in the test sample. It should be observed that from a set of calibration models for the spectra measured for all training temperatures (t_1, t_2, …,and t_K), the one with the temperature closest to that of the test sample is selected. Therefore, the most appropriate calibration model can be used to predict the concentration of the test sample. In addition ICS does not require the same subset of samples for all training temperatures. This characteristic of ICS offers some advantages over CPDS in process applications.

3. APPLICATION

An NIR data set was selected to compare the performance of CPDS and ICS. The data was collated from ninety five NIR spectra for 19 ternary mixtures of ethanol (C_1), water (C_2) and 2-propanol (C_3) recorded at five different temperature levels (30°C, 40°C, 50°C, 60°C and 70°C) in the range 580nm to 1049nm with a resolution of 1nm using an HP 8454 spectrophotometer equipped with a thermostable cell holder. The spectral region between 749nm and 849nm was used for slope and offset corrections and the 200 absorbances in the range 850nm to 1049nm were selected for the data analysis.

The 19 samples at each temperature were divided into 13 training and 6 test samples. Experimental details can be found in Wülfert, et al. (2000b), and the data is available at http://www-its.chem.uva.nl. The root-mean-squared error of prediction (RMSEP) was used as the performance criterion to assess the predictive power of ICS and CPDS. The RMSEP was calculated as follows:

$$RMSEP = \sqrt{\sum_{i=1}^{N_{test}} \frac{(y_{i,test} - \hat{y}_{i,test})^2}{N_{test}}} \qquad (8)$$

Here, $y_{i,test}$ is the known concentration of the *i*th test sample, $\hat{y}_{i,test}$ is the corresponding value estimated by ICS or the CPDS model. N_{test} denotes the number of test samples. The data analysis was undertaken on a Pentium computer using Matlab version 6.5 (Mathworks, Inc). All the programmes including ICS, CPDS and PLS1 were written in house.

4. RESULTS AND DISCUSSIONS

Mean-centred spectra ($\mathbf{X}$) were used as the prediction variables with the concentrations of water, ethanol and 2-propanol defining the predicted variables ($\mathbf{y}$) . For each chemical component, five local PLS1 calibration models with four latent variables (suggested by Wülfert, et al. 2000b) were built from the spectra of the training samples measured at five different temperatures (30 °C, 40°C, 50°C, 60°C and 70°C). The results of the local PLS1 models with 4 latent variables are provided for comparison in Table 1.

Table 1. RMSEP (×10^{-2} Mole fraction) of the test set for local models

Temp. (°C)		Components		
T_{pred}	T_{cal}	C_1	C_2	C_3
30	30	1.8	0.9	1.2
40	40	1.1	0.7	0.9
50	50	1.7	1.1	2.2
60	60	1.0	0.4	0.8
70	**70**	1.1	0.4	1.5

Before any detailed discussion on the performance of ICS is progressed, the validity of the assumption underpinning ICS is investigated. Fig. 1 shows the response patterns of all the three chemical species resolved by direct regression. It is evident that the variations in the response patterns with respect to temperature follow simple monotonic smooth non-linear functions.

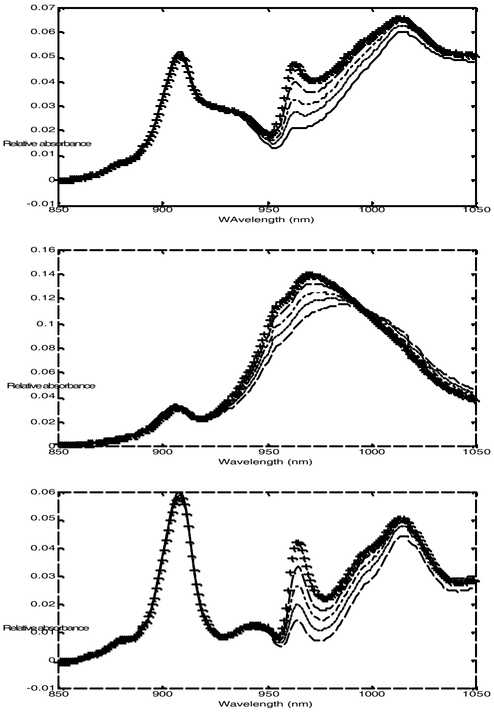

Fig.1. The resolved response patterns of ethanol (upper), water (middle) and 2-propanol (bottom) at 30°C (solid), 40°C (dotted), 50°C (dash dotted), 60°C (dash) and 70°C (plus).

The appropriateness of applying second order polynomials to model the temperature effects is also apparent from the agreement between the response patterns of the three chemical species at 50°C resolved by direct regression and the corresponding ones estimated by applying ICS using the spectra measured at temperatures other than 50°C (Fig. 2)

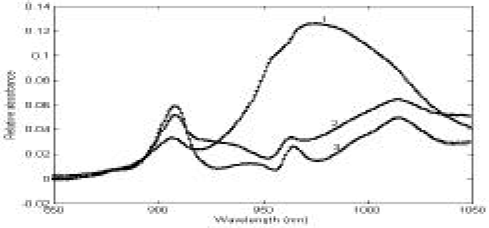

Fig.2. The response patterns of the three chemical species at 50 °C. Solid lines: resolved by direct regression; Dotted lines: estimated by ICS.

The previous results confirm the theoretical basis of ICS. The effectiveness of ICS for the correction of temperature induced spectral variations is illustrated through a typical set of results. Fig. 3 shows the 38 spectra for the 19 samples measured at 30°C and 50°C. Significant spectral differences introduced by temperature variations can be observed. When the ICS correction model built on the spectra of all the samples measured at training temperatures other than 50°C is applied, it can be observed that the temperature effects are effectively removed (Fig. 4).

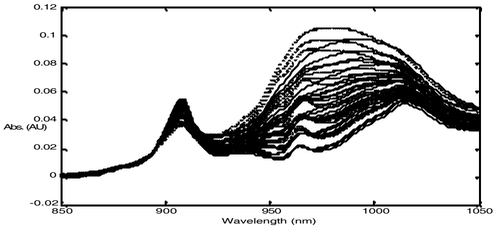

Fig.3. Spectra of all the 19 samples measured at 30°C (solid line) and 50°C (dotted line).

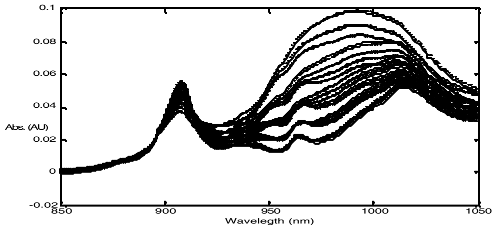

Fig.4. Spectra of all the 19 samples measured at 30°C (solid line) and those after corrections (dotted line) from 50°C to 30°C by ICS.

The predictive performance of the ICS correction model and those of CPDS are presented in Tables 2 and 3, respectively. The best results for the ICS and CPDS models for the test samples measured at 40°C, 50°C and 60°C are comparable to those for the local models. This demonstrates that both ICS and CPDS can remove, to some extent, the influence of the temperature-induced spectral variations on the predictions.

Table 2. RMSEP (×10^{-2} Mole fraction) of test sets after application of ICS correction model

T_{pred} (°C)	Comp.	T_{cal} (°C)			
		40	50	60	70
30	C_1	1.2	1.3	1.2	1.3
	C_2	2.3	2.3	2.7	2.8
	C_3	2.4	2.9	3.1	3.7
		30	50	60	70
40	C_1	1.4	1.2	1.0	0.8
	C_2	0.7	0.6	0.8	1.0
	C_3	1.5	1.1	1.3	1.6
		30	40	60	70
50	C_1	2.1	0.9	1.5	1.6
	C_2	1.8	1.3	1.3	1.4
	C_3	1.7	1.4	2.3	2.9
		30	40	50	70
60	C_1	1.8	1.2	1.1	0.9
	C_2	0.9	0.7	0.6	0.9
	C_3	1.8	1.4	1.1	1.1
		30	40	50	60
70	C_1	3.6	2.0	2.0	1.5
	C_2	5.9	4.6	3.6	2.3
	C_3	4.2	4.2	3.2	2.6

Table 3. RMSEP (×10^{-2} Mole fraction) of test sets after application of CPDS correction model

T_{pred} (°C)	Comp	T_{cal} (°C)			
		40	50	60	70
30	C_1	5.3	2.5	4.7	7.1
	C_2	2.6	1.2	1.1	1.2
	C_3	4.8	1.6	3.8	5.9
		30	50	60	70
40	C_1	1.2	1.3	1.5	1.7
	C_2	2.4	0.8	0.9	0.9
	C_3	2.5	1.6	1.8	2.1
		30	40	60	70
50	C_1	1.8	1.0	1.4	1.3
	C_2	1.8	1.4	1.0	0.7
	C_3	2.0	1.4	1.6	1.3
		30	40	50	70
60	C_1	1.7	1.1	1.0	3.4
	C_2	1.3	0.6	0.8	1.2
	C_3	1.4	1.2	0.7	4.0
		30	40	50	60
70	C_1	2.1	2.2	3.3	9.6
	C_2	3.3	2.6	1.7	3.6
	C_3	1.2	3.0	4.1	13

However the ICS and CPDS models demonstrate different abilities in terms of extrapolation. The average prediction errors of the CPDS models for the two extreme test temperatures, 30°C and 70°C, are significantly larger than those for ICS. In the CPDS models, with respect to different reference temperatures, variations in the prediction errors for water (C_2) for the test samples for the two extreme test temperatures are considerably smaller, compared with ethanol (C_1) and 2-propanol (C_3). This is a consequence of the higher absorption coefficients and larger temperature effects of water. The temperature effect of water dominates the variance in the spectra. CPDS tends to model the temperature effects of water. Hence when the CPDS models, with respect to different reference temperatures, are used for extrapolation, variations in the prediction errors of ethanol and 2-propanol will be larger than those for water. In contrast, ICS tries to model the temperature effects of each chemical species individually.

When ICS models, with respect to different reference temperatures, are used for extrapolation, the variations in the prediction errors of the three chemical species should be comparable. The results in Table 2 validate this hypothesis. Moreover, the results confirm that the calibration model built at the temperature closest to the test temperature provides better prediction results for the test samples after ICS correction, while CPDS does not possess such an advantage (Table 3).

5. CONCLUSION

It has been demonstrated that the influence of temperature-induced spectral variation on the predictive abilities of multivariate calibration models can effectively be removed through the proposed ICS approach. This is achieved by fitting polynomials for each element of the response patterns of the chemical species in mixtures against temperature. The estimated polynomials can then be used to predict the response patterns of the chemical species at the test temperature. The difference between the spectrum of the test sample measured at the test temperature and the corresponding spectrum theoretically measured at the reference temperature can then be calculated. After the temperature-induced spectral variations have been corrected, the calibration model can provide satisfactory predictions for test samples measured at temperatures other than the reference temperature.

Compared with CPDS, ICS has the following advantages: 1) Its implementation is straightforward. Unlike CPDS, only one parameter (the order of the polynomial) needs to be preset. In practice, the order of the polynomials can be set to a pre-defined vale of two. 2) ICS does not require the same training samples to be measured at different temperatures. 3) Since the calibration model built on the temperature closest to the test temperature can give more accurate predictions for test samples, it is straightforward to select the appropriate calibration model for the test samples.

A limiting factor in the application of ICS is that the system under study should be a white system (such as a crystallization process), i.e. there is no unknown chemical species in the system. This is not necessary for CPDS. Hence, CPDS is still the only chemometric method available for the correction of temperature fluctuations in grey systems, despite its limitations in terms of its application. Further research is focussing on standardization methods for the spectra of grey systems.

ACKNOWLEDGMENT

Dr Chen acknowledges the financial support of the EPSRC (GR/R19366/01 - KNOW-HOW and GR/R43853/01 – Chemicals Behaving Badly II).

REFERENCES

Blaser, W. W., R.A. Bredeweg, R.S. Harner, M.A. LaPack, A. Leugers, D.P. Martin, R.J. Pell, J. Workman Jr., and L.G. Wright (1995). Process analytical chemistry. *Anal. Chem.*, **67**, pp. 47R-70R.

Höskuldsson, A. (1988). PLS regression methods. *J. Chemom.*, **2**, pp. 211-228.

Iwata, T., J. Koshoubu, C. Jin and Y. Okubo (1997). Temperature dependence of the mid-Infrared OH spectral band in liquid water. *Appl. Spectrosc.*, **51**, pp. 1269-1275.

Lewiner, F., J.P. Klein, F. Puel and G. Févotte (2001). On-line ATR FTIR measurement of supersaturation during solution crystallization processes. calibration and applications on three solute/solvent. *Chemical Engineering Science*, **56**, pp. 2069-2084.

Swierenga, H., F. Wülfert, O.E. de Noord, A. P. Weijer, A.K. Smilde and L.M.C Buydens (2000). Development of robust calibration models in near infra-red spectrometric applications. *Anal. Chim. Acta*, **411,** pp. 121-135.

Wang, Y., D.J. Veltkamp and B.R. Kowalski (1991). Multivariate instrument standardization. *Anal. Chem.*, **63**, pp. 2750-2756.

Wang, Y. and B.R. Kowalski (1993). Temperature-compensating calibration transfer for near-infrared filter instruments. *Anal. Chem.*, **65**, pp. 1301-1303.

Wülfert, F., W. T. Kok, and A.K. Smilde (1998). Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Anal Chem.*, **70**, pp. 1761-1767.

Wülfert, F., W.T. Kok, O.E. de Noord and A.K. Smilde (2000a). Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. *Chemom. Intell. Lab. Syst.*, **51**, pp. 189-200.

Wülfert, F., W.T. Kok, O.E. de Noord and A.K. Smilde (2000b). Correction of temperature-induced spectral variation by continuous piecewise direct standardization. *Anal. Chem.*, **72**, pp. 1639-1644.