

Calibration Transfer by Generalized Least Squares

Barry M. Wise, Harald Martens, Martin Høy, Rasmus Bro & Per B. Brockhoff

Abstract

The instrument standardization/calibration transfer problem has been addressed by a wide variety of methods. In this work, we investigate the ability of Generalized Least Squares (GLS) preprocessing methods to deal with artifacts caused by changes in spectroscopic instrumentation that would normally require the development of completely new calibration models. GLS preprocessing works by measuring a number of transfer samples on two instruments (or one instrument at separate times). These samples can be used to estimate an offset and shift in the covariance structure of the data due to instrument differences. The method of GLS preprocessing shown recently by Harald Martens *et. al.* can then be used to remove variation in the data which is not common to both instruments. Calibration models can then be built on data from one of the instruments and used on the other, with the GLS preprocessing applied prior to predictions on new samples. The GLS method is tested on two data sets from near infrared spectroscopy, one being pseudo-gasoline mixtures and the other corn measured on three instruments. Comparisons are made to other calibration transfer methods including Piece-wise Direct Standardization (PDS) and Orthogonal Signal Correction (OSC).

Preprocessing by Generalized Least Squares

Martens *et. al.* [1] recently showed how the problem of modelling *undesired* or *irrelevant* but known error covariances can be mitigated by multiplying the input data by the square root of the known error covariance matrix, thereby *shrinking* the data space in the chosen off-axis directions. This amounts to making Generalized Least Squares (GLS) methods out of Ordinary Least Squares (OLS) methods, without having to change the OLS method itself.

For instrument standardization applications, the difference between instruments can be modelled as an offset (additive background) term and a covariance defined by the differences between samples measured on both instruments. Thus the response of two instruments can be made more similar by adjusting for the offset and multiplying each of the instrument responses by the square root of the estimated difference covariance. This has the effect of shrinking directions in the data space which are not common to both of the instruments.

Development of the GLS Weighting Matrix

For standardization applications, a weighting matrix **W** will be developed that will be applied to both the standard instrument and the instrument to be standardized. Given a set of transfer samples **X**₁ and **X**₂ that have been measured on two instruments, estimate the offset term as the difference between the mean of the instruments:

$$\bar{\mathbf{x}}_d = \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$$

The mean adjusted difference between the samples is then

$$\mathbf{X}_d = \mathbf{X}_1 - \mathbf{X}_2 + \mathbf{1}\mathbf{x}_d$$

where **1** is a column vector of ones of appropriate dimension. The covariance of the difference is then estimated in the usual way except that a small value (typically $\sim 1e-6$) is added to the diagonal:

$$\mathbf{C}_d = \mathbf{X}_d^T \mathbf{X}_d / (m-1) + \mathbf{I}$$

where *m* is the number of transfer samples. The constant is added to assure that the covariance matrix is of full rank, and is the sole adjustable parameter in the method. **C**_{*d*} is then decomposed using either a singular value (or eigenvector) decomposition as:

$$\mathbf{USV}^T = \mathbf{C}_d$$

The eigenvalues (diagonal of **S**) are then adjusted to have a mean of 1 and the square root taken to form the matrix of adjusted singular values

$$\mathbf{S}_{adj} = \text{sqrt}(\mathbf{S} * n / \text{trace}(\mathbf{S}))$$

where *n* is the number of variables. Finally, the weighting matrix is formed by

$$\mathbf{W} = \mathbf{VS}_{adj}^+ \mathbf{V}^T$$

where **S**_{*adj*}⁺ is the pseudo-inverse of **S**_{*adj*} where care has been taken not to invert singular values near zero.

Application of the GLS Weighting and Offset Term

The mean centered standard instrument data can now be preprocessed by post multiplying it by the weight **W**:

$$\mathbf{X}_{1adj} = (\mathbf{X}_1 - \mathbf{1}\mathbf{x}_1) \mathbf{W}$$

Calibration of instrument 1 then proceeds as normal, using the regression method of choice. Once the regression model has been developed, it can be applied to the adjusted instrument 2 data, defined as

$$\mathbf{X}_{2adj} = (\mathbf{X}_2 - \mathbf{1}\mathbf{x}_2 + \mathbf{1}\mathbf{x}_d) \mathbf{W}$$

Reference Methods

The GLS method described here is compared with the Piece-wise Direct Standardization (PDS) and Orthogonal Signal Correction (OSC). The PDS method [2,3] attempts to find a transform that essentially predicts the response of the standard instrument from the response of second instrument and has found wide use in calibration transfer problems.

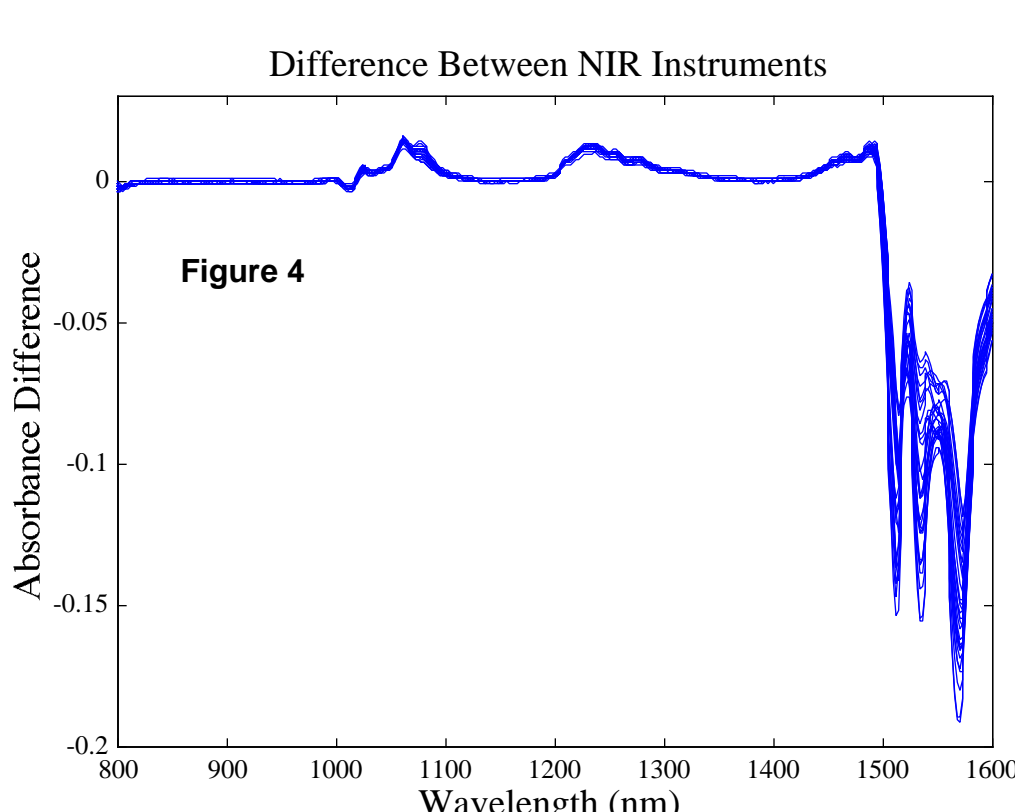
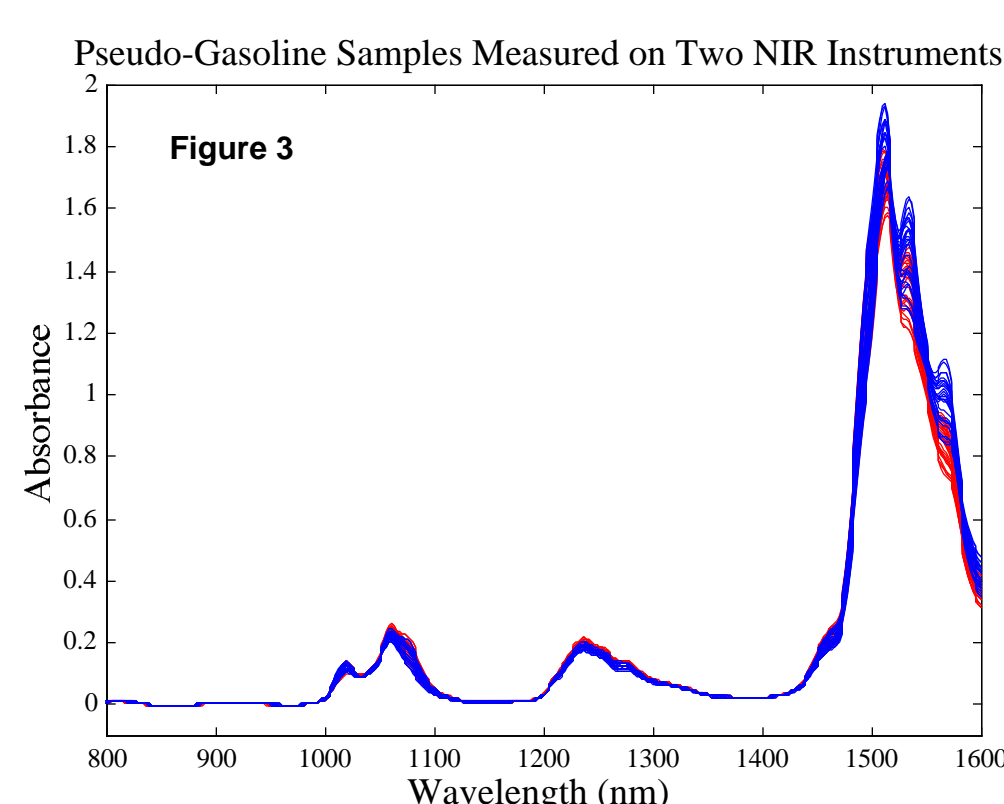
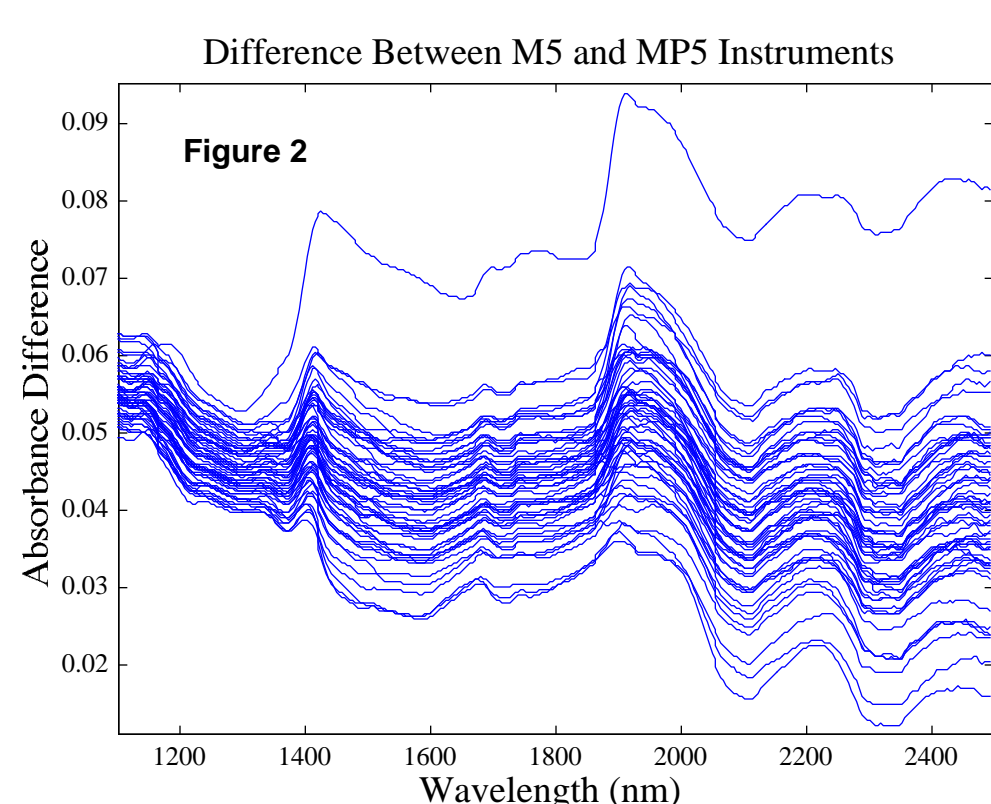
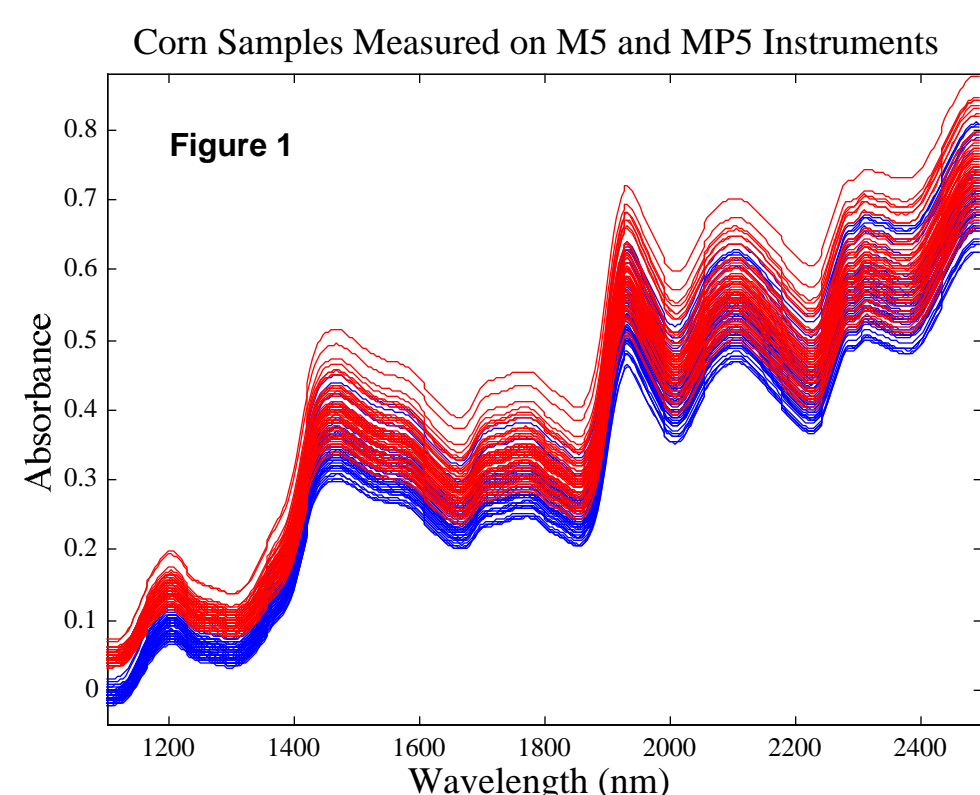
OSC, often used as a preprocessing method for single calibrations, can also be an effective calibration transfer method. In this application, the OSC model is developed with transfer samples from both instruments and is then applied to data from both instruments. The general idea is that variance that is unique to each instrument and not correlated with the property of interest is removed.

GLS is perhaps more similar to OSC in that it attempts to remove the difference between instruments rather than making one instrument look like the other. Like PDS, however, GLS does not require property (**Y**) values, which OSC does.

Available Data

A data set consisting of 80 corn samples measured on three instruments (M5, MP5 and MP6) was provided by Mike Blackburn at Cargill. Reference values for moisture, oil, protein and starch were available for all the samples. An example of the samples measured on two instruments is shown in Figure 1, and the difference between the samples is shown in Figure 2.

A second data set consisted of 30 samples of a 5 component pseudo-gasoline mixture measured on two instruments. This data set was developed at AMOCO by Ernie Baughman. Spectra of the 30 samples measured on each of the instruments is shown in Figure 3. The differences between samples are shown in Figure 4.



Selection of Transfer Samples

Transfer samples can be selected in a number of different ways. Experience with PDS shows that the best transfers are achieved when the samples are selected based upon their leverage in the calibration model of the analyte of interest. OSC appears to work best when the transfer samples are chosen based on their leverage in a PCA model of the data. GLS standardization also appears to work best when the transfer samples are chosen this way.

Experimental-Corn Data

The original 80 sample data set was split into 60 samples for training purposes and 20 samples for testing. PLS calibrations were developed for each instrument-analyte pair. These calibrations were then applied to the test samples on all instruments and errors calculated. The results are shown in the first three lines of Table 1 and summarized in Figure 6. The average root-mean-square error of prediction (RMSEP) for all analytes where the model was built on the same instrument as the prediction set was 0.1706. When the model was built on a different instrument than the prediction set the mean RMSEP was 1.0052, an approximately 6-fold increase in prediction error, as might be expected.

PDS models were developed for each combination of instruments and analyte. All of the PDS transforms were based on a window width of 11 channels and 1 factor PCR submodels. As shown in Table 1, the predictions were much improved compared to no standardization, and suffered only moderately compared to same model-instrument predictions, with a mean RMSEP of 0.2289.

GLS models were developed over all instrument pairs. Because the transfer sample selection and GLS method itself is independent of analyte, separate models for each analyte were not required, unlike both PDS and OSC. PLS models were then developed individually for each instrument pairing and analyte. Typical calibration and prediction curves for GLS standardization are shown in Figure 5, which is for standardizing MP5 to M5 for the moisture data. Results of the GLS standardization are shown in Table 1. The average RMSEP of 0.1831 is much improved compared to PDS, and nearly equal to the same model-instrument predictions of 0.1706. A search for the minimum RMSEP over 5-8 latent variables (LVs) was also performed. This is a best case scenario, the results that would be obtained if the number of LVs chosen was optimal in all cases. The result was a mean RMSEP of 0.1662.

OSC models were developed for each combination of instruments and analyte. Due to the difficulty in estimating optimal numbers of OSC and PLS factors with cross-validation, the optimal cases are given over the range of 1-3 OSC components and 3-8 LVs. The mean RMSEP was 0.1637. Again, this is a best case scenario. The single best combination of OSC and PLS components (3 and 5, respectively) is also shown, with a mean RMSEP of 0.2328.

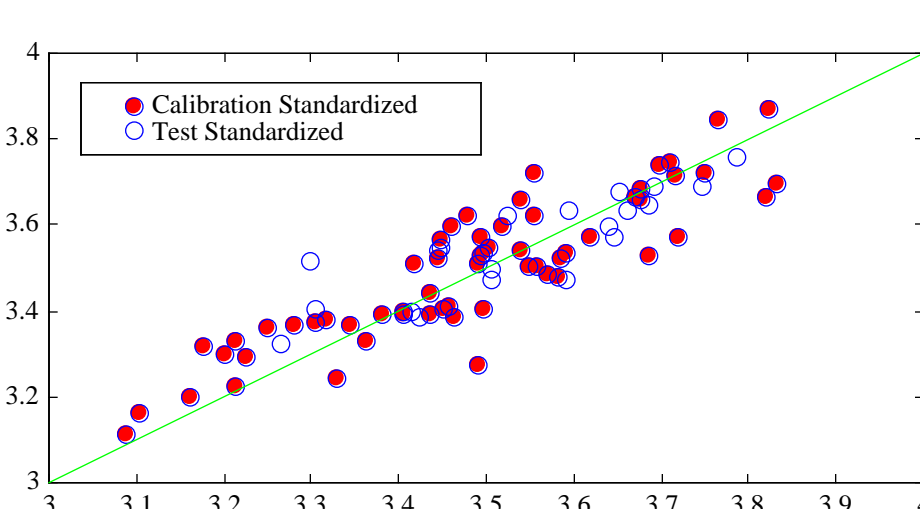
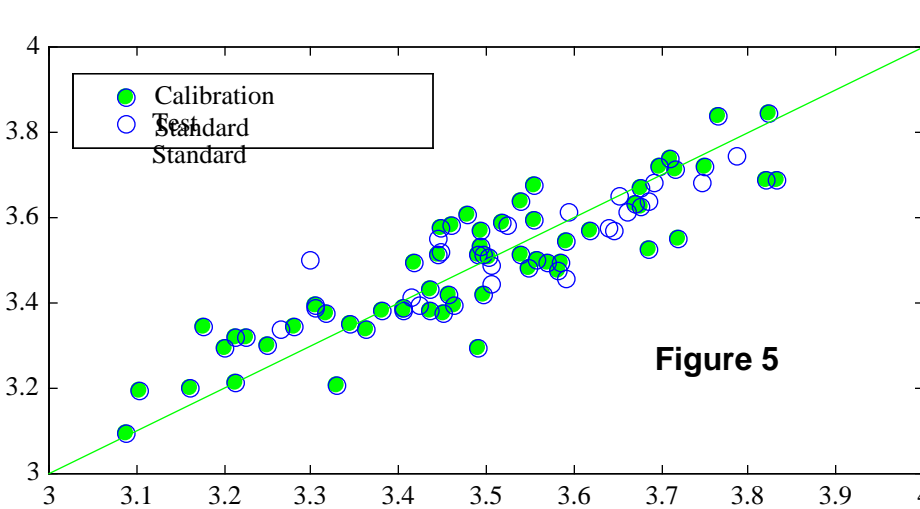
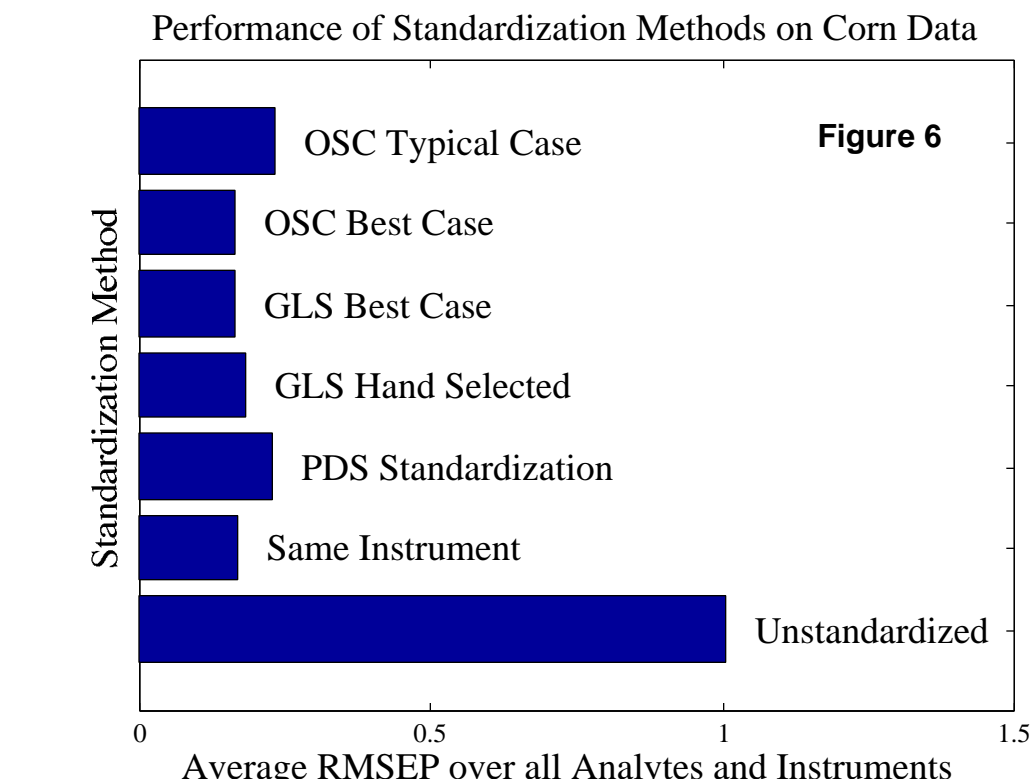


Table 1. Summary of Corn Data Standardization Results
(values given are RMSEP)

Prediction Instrument

	Moisture			Oil			Protein			Starch			Average
	M5	MP5	MP6	M5	MP5	MP6	M5	MP5	MP6	M5	MP5	MP6	
Preds													
M5	0.0187	1.4166	1.5123	0.0361	0.1274	0.1568	0.1302	1.2685	1.3241	0.2077	2.0949	1.6601	
MP5	1.1693	0.1460	0.3547	0.2751	0.0885	0.1516	1.2719	0.1720	0.2782	3.5674	0.4091	0.6119	1.0052
MP6	1.0921	0.2849	0.1667	0.3148	0.1926	0.0819	0.8982	0.2403	0.1876	3.1865	0.6754	0.4031	0.1706
PDS Standardization													
M5	-	0.3951	0.4671	-	0.0932	0.0755	-	0.1699	0.1849	-	0.3362	0.3710	
MP5	0.2342	-	0.1749	0.0876	-	0.0944	0.1401	-	0.1880	0.3455	-	0.3972	0.2289
MP6	0.2068	0.1601	-	0.0920	0.1035	-	0.1553	0.1770	-	0.4147	0.4290	-	
GLS Standardization, LVs hand selected													
M5	-	0.1592	0.1908	-	0.0859	0.0952	-	0.1531	0.1679	-	0.3314	0.3420	
MP5	0.1391	-	0.1477	0.0479	-	0.0770	0.1722	-	0.2110	0.2830	-	0.4381	0.1831
MP6	0.1990	0.1521	-	0.0603	0.0816	-	0.1687	0.1570	-	0.1873	0.3473	-	
GLS Standardization, best over 5-8 LVs													
M5	-	0.1545	0.1897	-	0.0688	0.0783	-	0.1485	0.1602	-	0.3039	0.3350	
MP5	0.1248	-	0.1258	0.0479	-	0.0696	0.1448	-	0.1721	0.2405	-	0.3709	0.1662
MP6	0.1902	0.1177	-	0.0590	0.0753	-	0.1358	0.1570	-	0.1873	0.3316	-	
OSC Standardization, best over all cases, 1-3 OSC, 3-8 LVs													
M5	-	0.1630	0.1733	-	0.0816	0.0710	-	0.1433	0.1502	-	0.3002	0.3293	
MP5	0.1945	-	0.1580	0.0710	-	0.0739	0.1394	-	0.1988	0.2640	-	0.4259	0.1637
MP6	0.1466	0.1320	-	0.0607	0.0686	-	0.1568	0.1449	-	0.2253	0.3744	-	
OSC Standardization, best single case, 3 OSC 5 LVs													
M5	-	0.2218	0.2611	-	0.0835	0.0742	-	0.1588	0.1502	-	0.3250	0.3515	
MP5	0.3097	-	0.2176	0.0830	-	0.0834	0.1601	-	0.2388	0.4206	-	0.4506	0.2328
MP6	0.3299	0.1379	-	0.0826	0.1157	-	0.1684	0.2154	-	0.5119	0.4363	-	



Experimental-Pseudo-gasoline Data

The 30 samples in the pseudo-gasoline data were split into a calibration set of 20 samples and a test set of 10 samples. Care was taken to not leave out unique or high leverage samples. Individual PLS models were developed for each of the 5 analytes in the data. Because of the small size of the data calibration set, the number of LVs in the models was fixed at 5 based on the fact that it is a known 5 component system.

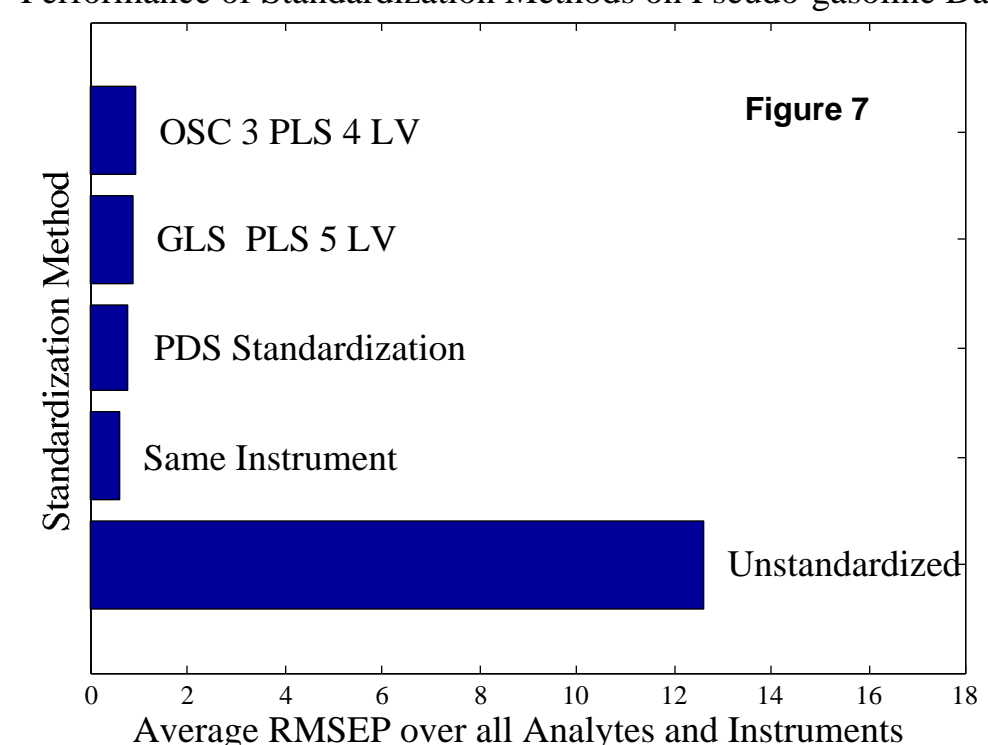
The average RMSEP over all analytes and for both instruments for the test set as measured on the same instrument was 0.5836. For the unstandardized second instruments the RMSEP increased to 12.63. Results are compared in Figure 7.

Individual PDS models were built for each combination of instruments and analytes. The PDS models all used a window of 3 channels. The mean RMSEP for PDS standardization was 0.7955, approximately 36% higher than using the same instrument, but drastically improved compared to no standardization.

GLS weighting matrices were developed for each of the two instruments. PLS models with 5 factors were found to be optimal for the GLS preprocessed data. The mean RMSEP using GLS preprocessing was 0.8823.

The optimal OSC combination for prediction of new samples was 3 OSC components and 4 PLS LVs. Using this combination, the RMSEP was 0.9437.

Performance of Standardization Methods on Pseudo-gasoline Data



Discussion

All of the standardization methods used here, PDS, GLS and OSC, performed well. In the case of the corn data, the GLS and OSC methods produced predictions from standardized instruments and new samples with accuracy very close to that of the original instruments. In the pseudo-gasoline data, PDS proved best, with GLS close behind.

Given the similarity of the results, issues concerning usability should be considered. The GLS preprocessing method is particularly easy to use because it has few adjustable parameters (only 5, to which it is fairly insensitive) and requires only one model per instrument, since it is independent of analyte. GLS does, however, require that new calibration models be built on the standard instrument after GLS preprocessing.

The PDS method has several adjustable parameters, the most important of which is the window width. This is typically easy to set with cross-validation, even in a small transfer set (5-10 samples). Other parameters involve the development of the individual sub-model, however, in this work, the default PLS_Toolbox [6] values for these parameters were used.

Of the methods tested, OSC is probably the most difficult to use due to difficulties with establishing the proper number of OSC components and PLS components in the subsequent models. Once OSC is performed on a data set, cross-validation results are often unreliable. For this reason, the OSC step should be included in the cross-validation, however, it is not clear how to best do this in the calibration transfer situation. It is for these reason that many combinations of OSC and PLS components were tried and the best of these reported. In practice, without additional test sets, this would not in general be possible.

Conclusion

GLS appears to be quite competitive with PDS and OSC for standardization purposes. The method is easy to apply, requires only one selection of transfer samples and model for each instrument pair, and has few adjustable parameters.

References

- [1] H. Martens, M. Høy, B.M. Wise, R. Bro and P.B. Brockhoff, "GLS Preprocessing of Multivariate Data," submitted to *J. Chemometrics*, May 2001.
- [2] Y. Wang, D.J. Veltkamp and B.R. Kowalski, "Multivariate Instrument Standardization," *Anal. Chem.*, **63**(23), pps 2750-2756, 1991.
- [3] Z. Wang, T. Dean and B.R. Kowalski, "Additive Background Correction in Multivariate Instrument Standardization," *Anal. Chem.*, **67**(14), pps 249-260, 1995.
- [4] S. Wold, H. Antti, F. Lindgren and J. Öhman, "Orthogonal Signal Correction of Near-Infrared Spectra," *Chemo. and Intell. Lab. Sys.*, **44**, pps 175-185, 1998.
- [5] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg and S. Wold, "An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra," *Chemo. and Intell. Lab. Sys.*, **44**, pps 229-244, 1998.
- [6] B.M. Wise and N.B. Gallagher, *PLS_Toolbox 2.1 for use with MATLAB*, Eigenvector Research, Inc., Manson, WA USA, 2000.

