Received: 29 May 2020 | Revised: 29 September 2020 | Accepted: 29 October 2020

DOI: 10.1002/cem.3319

RESEARCH ARTICLE

Journal of CHEMOMETRICS WILEY

Graph-based calibration transfer

Ramin Nikzad-Langerodi | **Florian Sobieczky**

Data Science Group, Software Competence Center Hagenberg (SCCH), Hagenberg, 4232, Austria

Correspondence
Ramin Nikzad-Langerodi, Data Science Group, Software Competence Center Hagenberg (SCCH), Hagenberg 4232, Austria.
Email: ramin.nikzad-langerodi@scch.com

Funding information
FFG

Abstract

The problem of transferring calibrations from a primary to a secondary instrument, that is, calibration transfer (CT), has been a matter of considerable research in chemometrics over the past decades. Current state-of-the-art (SoA) methods like (piecewise) direct standardization perform well when suitable transfer standards are available. However, stable calibration standards that share similar (spectral) features with the calibration samples are not always available. Towards enabling CT with arbitrary calibration standards, we propose a novel CT technique that employs manifold regularization of the partial least squares (PLS) objective. In particular, our method enforces that calibration standards, measured on primary and secondary instruments, have (nearly) invariant projections in the latent variable space of the primary calibration model. Thereby, our approach implicitly removes interdevice variation in the predictive directions of **X**, which is in contrast to most SoA techniques that employ explicit preprocessing of the input data. We test our approach on the well-known corn benchmark data set employing the National Bureau of Standards and Technology (NIST) glass standard spectra for instrument standardization and compare the results with current SoA methods.

KEYWORDS

calibration transfer, graph Laplacian, manifold regularization, partial least squares, transfer learning

1 | INTRODUCTION

Calibration transfer (CT), sometimes referred to as instrument standardization in chemometrics, is the process of transferring a calibration model from one instrument to another.[1-3] Ideally, CT preserves the accuracy and precision of a calibration model developed on a primary instrument, that is, providing statistically identical analysis of the same samples measured on the secondary instrument. Historically, CT has been addressed by (i) model updating or (ii) measuring a set of so-called calibration standards on both instruments in order to derive a correction for the difference in the instrumental response. The former can be considered more generic and copes with any type of change related to the measurement condition such as environmental influences, matrix effects, or instrumental changes. Slope and bias correction,[4] calibration set augmentation,[5] model updating via Tihkonov regularization,[6] standard-free CT,[7-9] and domain-invariant modeling[10-12] all belong to this category and have been applied with success to CT problems. However, these methods usually require a considerable amount of (additional) samples (and reference measurements), and deciding between maintenance and recalibration is not always straightforward. CT by means of calibration standards, on the other hand, solely requires a small set of samples that can be measured on both devices and does not require any additional reference values. In this second category, direct (DS) and piecewise direct standardization (PDS)

 © 2020 John Wiley & Sons, Ltd.

can be considered the gold standard.[13] Both operate by learning a multivariate transformation (mapping) such that the instrumental response of the secondary instrument matches with the one of the primary instrument. Several alternative techniques have been recently proposed such as generalized least squares weighting (GLSW), which downweighs directions in the predictor matrix that are associated with large between-device differences;[14,15] orthogonal signal correction (OSC) and related techniques based on orthogonalization of the regression vector to differences between devices;[16,17] spectral subspace transform (SST), which reconstructs the secondary instruments' response in the space spanned by the primary instrument's principal components;[18] and spectral regression (SR) and PLSCT, which perform DS between (nonlinear) embeddings of the calibration standards[19,20] and data integration methods such as Joint and Unique Multiblock Analysis (JUMBA)[21] or Joint-Y partial least squares (PLS)[22] that aim at deriving models based on latent variables (LVs) that are common to primary and secondary instrumental responses.

Looking at the vast majority of publications that have addressed CT over the past decades, two general observations can be made: Most techniques seem to work well when the standardization set is a (carefully selected) subset of the calibration set, that is, if samples and standards share similar spectral features. In addition, most CT methods proposed so far perform explicit corrections of the instrumental response rather then modeling the interdevice variation implicitly.

Towards CT with arbitrary calibration standards, we thus propose a method based on a novel, regularized PLS variant that implicitly corrects for interdevice variation in the directions of the input matrix that are related to the target property. In particular, our approach aims at preserving the special graph in which only the matched calibration standards are connected by an edge while finding an embedding of the (primary) calibration samples that is informative with respect to the target property. Along this line, our approach can be considered a special case of previous work on locality preserving projection[23] in the context of PLS modeling.[24,25]

2 | THEORY

2.1 | Notation

We follow standard notation in chemometrics, where upper and lower case boldface symbols denote matrices and vectors, respectively, and scalars are denoted by nonboldface symbols. Unless otherwise stated, vectors are column vectors, and $^{\mathrm{T}}$ and $^{-1}$ denote the transpose and inverse operation, respectively. By $\mathbf{I}$ and $\mathbf{0}$, we denote the identity and null matrix (with appropriate dimensions), respectively, and by $\otimes$ the tensor (Kronecker) product between two matrices. Without loss of generality, we assume that the response y is univariate. The matrices $\mathbf{A}$, $\mathbf{D}$, and $\mathbf{L}$ are used to denote the adjacency, degree, and Laplace matrix of a simple, undirected graph, and $|\cdot|_{\mathrm{F}}$ denotes the Frobenius norm. Comma and semicolon notations are used to denote horizontal and vertical stacking of scalars and matrices, for example, $[\mathbf{x}, y]$ and $[\mathbf{x}; \mathbf{y}]$.

2.2 | Problem formulation

We describe our method by taking the situation where there are N labeled calibration samples from the primary instrument, that is, $\{(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R} | i = 1, ..., N\}$, and a set of K calibration standard samples $\left\{\left(\mathbf{x}_i^{(p)}, \mathbf{x}_i^{(s)}\right) \in \mathbb{R}^d \times \mathbb{R}^d | i = 1, ..., K\right\}$, measured on the primary and secondary instrument. It is assumed that the same d variables are measured on both primary and secondary instrument, that is, that the wavelength axis is stable across the instruments. The goal of CT is to derive a model $h : \mathbb{R}^d \to \mathbb{R}$ from the primary calibration and the calibration standard samples that has a small root mean squared error of prediction (RMSEP) when applied to samples measured on the secondary instrument.

2.3 | Graph-based CT

In order to minimize interdevice variation when building the primary calibration model, we proceed as follows: We let $\mathbf{X}_p = \left[\mathbf{x}_1^{(p)}, \mathbf{x}_2^{(p)}, ..., \mathbf{x}_K^{(p)}\right]^{\mathrm{T}}$ and $\mathbf{X}_s = \left[\mathbf{x}_1^{(s)}, \mathbf{x}_2^{(s)}, ..., \mathbf{x}_K^{(s)}\right]^{\mathrm{T}}$ be matrices holding the matched calibration standards measured on the primary and the secondary instrument, respectively. We further let $\mathbf{K} = [\mathbf{X}_p; \mathbf{X}_s] = [\mathbf{k}_1, ..., \mathbf{k}_{2K}]^T$; $\mathbf{X}$ and $\mathbf{y}$ hold the calibration samples from the primary instrument, and

$$\begin{aligned} J &= \frac{1}{2}\sum_{i=1}^{2K}\sum_{j=1}^{2K}\left(\mathbf{k}_i^{\mathrm{T}}\mathbf{w}-\mathbf{k}_j^{\mathrm{T}}\mathbf{w}\right)^2\mathbf{A}_{i,j} \\ &= \mathbf{w}^{\mathrm{T}}\left[\sum_{i=1}^{2K}\mathbf{k}_i\mathbf{D}_{i,i}\mathbf{k}_i^{\mathrm{T}}-\sum_{i=1}^{2K}\sum_{j=1}^{2K}\mathbf{k}_i\mathbf{A}_{i,j}\mathbf{k}_j^{\mathrm{T}}\right]\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{D}-\mathbf{A})\mathbf{K}\mathbf{w} \\ &= \mathbf{w}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{L}\mathbf{K}\mathbf{w}, \end{aligned} \tag{1}$$

with $\mathbf{D}=\mathbf{I}$, $\mathbf{A}=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\otimes\mathbf{I}$, $\mathbf{L}=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}\otimes\mathbf{I}$, and $\mathbf{I}$ denoting the $K\times K$ identity matrix. We then compute

$$\mathbf{w}^* := \left[\frac{\mathbf{y}^{\mathrm{T}}\mathbf{X}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\left(\mathbf{I}+\frac{\gamma}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\boldsymbol{\Gamma}\right)^{-1}\right]^{\mathrm{T}}, \tag{2}$$

with $\boldsymbol{\Gamma}=\mathbf{K}^{\mathrm{T}}\mathbf{L}\mathbf{K}$, which is the minimizer of (see Appendix A for derivation)

$$\min_{\mathbf{w}}||\mathbf{X}-\mathbf{y}\mathbf{w}^{\mathrm{T}}||_{\mathrm{F}}^2+\gamma J, \tag{3}$$

where γ is a regularization parameter that controls the trade-off between attaining a subspace that is predictive with respect to the response $\mathbf{y}$ of the (primary) calibration samples and minimizing the Euclidean distance between the matched CT samples in the LV space. The first term in Equation (3) maximizes the covariance between $\mathbf{X}$ and $\mathbf{y}$ and is thus equivalent to the PLS objective. The second term corresponds to a so-called manifold regularization term and biases the PLS weight vector $\mathbf{w}$ such that the interdevice variation becomes smaller. After computing $\mathbf{w}^*$, we proceed in analogy with the well-known NIPALS algorithm; that is, we perform following iterations:

1. Mean centering $\mathbf{X}=\mathbf{X}_0\left(\mathbf{I}-\frac{1}{N}\mathbf{1}\mathbf{1}^{\mathrm{T}}\right);\mathbf{y}=\mathbf{y}_0-\bar{\mathbf{y}}$, where $\mathbf{X}_0$ and $\mathbf{y}_0$ refer to the raw data
2. Stack (deflated) calibration standard matrices $\mathbf{K}=[\mathbf{X}_p;\mathbf{X}_s]$
3. Compute regularization term $\boldsymbol{\Gamma}=\mathbf{K}^{\mathrm{T}}\mathbf{L}\mathbf{K}$
4. Compute weight vector $\mathbf{w}^*=\left[\frac{\mathbf{y}^{\mathrm{T}}\mathbf{X}}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\left(\mathbf{I}+\frac{\gamma}{\mathbf{y}^{\mathrm{T}}\mathbf{y}}\boldsymbol{\Gamma}\right)^{-1}\right]^{\mathrm{T}}$
5. Normalize $\mathbf{w}^*=\frac{\mathbf{w}^*}{|\mathbf{w}^*|}$
6. Compute of scores $\mathbf{t}=\mathbf{X}\mathbf{w}^*;\mathbf{t}_p=\mathbf{X}_p\mathbf{w}^*;\mathbf{t}_s=\mathbf{X}_s\mathbf{w}^*$
7. Regression $c=\frac{\mathbf{t}^{\mathrm{T}}\mathbf{y}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}}$
8. Compute loadings $\mathbf{p}=\frac{\mathbf{X}^{\mathrm{T}}\mathbf{t}}{\mathbf{t}^{\mathrm{T}}\mathbf{t}};\mathbf{p}_p=\frac{\mathbf{X}_p^{\mathrm{T}}\mathbf{t}_p}{\mathbf{t}_p^{\mathrm{T}}\mathbf{t}_p};\mathbf{p}_s=\frac{\mathbf{X}_s^{\mathrm{T}}\mathbf{t}_s}{\mathbf{t}_s^{\mathrm{T}}\mathbf{t}_s}$
9. Deflation: Transform $\mathbf{X}$ into $\mathbf{X}-\mathbf{t}\mathbf{p}^{\mathrm{T}}$, $\mathbf{X}_p$ into $\mathbf{X}_p-\mathbf{t}_p\mathbf{p}_p^{\mathrm{T}}$, and $\mathbf{X}_s$ into $\mathbf{X}_s-\mathbf{t}_s\mathbf{p}_s^{\mathrm{T}}$
10. Aggregation: Transform $\mathbf{W}$ into $[\mathbf{W},\mathbf{w}^*]$, $\mathbf{p}$ into $[\mathbf{p},\mathbf{p}]$, and $\mathbf{c}$ into $[\mathbf{c};c]$
11. Return to 2 until A LVs have been computed
12. Compute regression coefficients $\mathbf{b}=\mathbf{W}(\mathbf{p}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{c}$

2.4 | Discussion

As an illustrative example, consider the situation of having exactly two calibration standard samples (Figure 1). For the adjacency matrix $\mathbf{A}$ in Equation (1), we then have

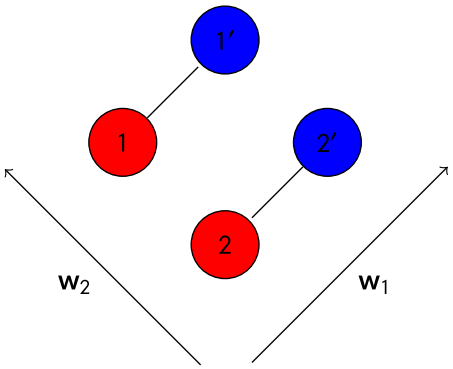

FIGURE 1 Illustrative example of the idea of graph-based calibration transfer (GCT) with two calibration transfer samples measured on the primary (red) and secondary (blue) instrument. Only matched samples (i.e., the same sample measured on both instruments) have a vertex. Projection on direction $\mathbf{w}_1$ and $\mathbf{w}_2$ results in large and small interdevice difference in the latent variable (LV) space

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{0} & \mathbf{I} \\ \mathbf{I} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix};$$

that is, only matched standard samples have a vertex in the corresponding graph defined over $\mathbf{K}$. Consequently, the regularization term in Equation (3) is large if the Euclidean distances btw. the projections of the matched standard samples are large. Thus, increasing the magnitude of the parameter γ reduces the difference between primary and secondary instruments in the y-predictive direction of $\mathbf{X}$. Since the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is always positive semi-definite, the regularizer $\mathbf{w}^{\mathrm{T}}\boldsymbol{\Gamma}\mathbf{w}$ is convex and preserves the convexity of the objective function.[25] Thus, the unique solution of Equation (3) is attained as the root of the derivative and has the closed-form solution of Equation (2).

3 | EXPERIMENTAL SECTION

3.1 | Data set

All experiments were carried out using the Cargill corn benchmark data set obtained from Eigenvector Research Inc.[1] The data set comprises near-infrared (NIR) spectra from a set of 80 samples of corn measured on three different spectrometers at 700 spectral channels in the wavelength range 1100–2498 nm at 2-nm intervals.

3.2 | Evaluation

All experiments were conducted using in-house code implemented in Python programming language. Calibration (60) and validation (20) sets were derived by means of the Kennard–Stone algorithm[26] from the primary instrument samples. The spectra of the first three National Bureau of Standards and Technology (NIST) glass standards measured on the primary and secondary instrument were used together with the calibration set of the primary instrument in order to fit calibration models that were subsequently validated using the validation samples from the secondary instrument. The calibration set was mean centered, and the validation set recentered to the mean of the calibration set. Unless explicitly stated, spectra of the calibration standards were not mean centered. Pair-wise CT between instruments m5, mp5, and mp6 for the prediction of moisture, oil, protein, and starch were considered. For experiments involving NIST standards, we compared our approach with (primary) PLS models established on the primary instrument's calibration samples and validated on the secondary instrument's validation samples and (secondary) PLS models established and validated using secondary calibration and validation samples, respectively. For experiments involving corn samples as calibration standards, 10 sample spectra were selected from the primary calibration samples by means of the Kennard–Stone algorithm and used along with the corresponding spectra from the secondary instrument for CT. Comparison of our approach with DS was carried out by correcting the calibration samples (from the primary instruments) s.t.

[1]https://www.eigenvector.com/data/Corn/ (accessed 11 April 2018).

$$\widetilde{\vec{X}}_{\text{cal}} = \vec{X}_{\text{cal}}\mathbf{F}, \tag{4}$$

before model building with $\mathbf{F} = \mathbf{X}_p^{\dagger}\mathbf{X}_s$, and $\mathbf{X}_p^{\dagger}$ is the pseudo-inverse of the calibration standard spectra from the primary instrument. Similarly, PDS transforms were obtained based on a window width of 11 channels and one factor PLS models.

4 | RESULTS

In most works that have been published on CT so far, a subset of the calibration samples constitutes the transfer standards. However, this requires that the samples are sufficiently stable, because physical or chemical degradation between measurement on the primary and secondary instrument would lead to deterioration of CT efficiency. For a large variety of sample types, for example, agricultural or related (food) products, stability is not given. Ideally, CT should therefore be undertaken with stable materials that can have different spectral features compared with the calibration samples. However, standard CT techniques usually break down in such scenarios. As an example, Figure 2 shows application of DS of the corn spectra from instrument m5 using the NIST glass standards m5 NIST and mp6 NIST. Obviously, the difference in the spectral response on the corn samples can not be corrected for by the glass samples, because the spectral features of the latter are introduced in the standardization. Thus, a calibration based on the standardized m5 spectra (i.e., m5F) will not generalize well to samples analyzed on instrument mp6 (note that the NIST standards were originally intended to either test or align the instrument output function for wavelength or photometric accuracy rather than for instrument standardization).

4.1 | Proof of principle

In order to avoid that the spectral features of the calibration standards are carried over to the spectra of the calibration samples, our approach standardizes the instrumental response in the y-predictive LV space of the primary calibration model. However, rather than employing DS between the CT samples in the LV space of the primary calibration model as proposed by Zhao and coworkers,[20] our approach aims at implicit standardization by jointly mapping the (primary) calibration samples and calibration standards measured on the primary and secondary instrument such that (i) the LVs are predictive with respect to the response and (ii) the distance between the projections of matched calibration standard samples is small. Figure 3 shows the projections of the NIST glass standards measured on m5 and mp6 together with the calibration set from m5 on the first 2 LVs for an increasing amount of regularization. Without regularization, the LV space corresponds to the ordinary PLS subspace in which m5 NIST and mp6 NIST samples differ mostly along the first LV. With increasing γ, the difference between the calibration standards becomes smaller. Notably, the reconstructed spectra of the calibration samples (measured on m5) and validation samples (measured on mp6) from the corresponding LVs become increasingly similar with increasing regularization (Figure 3, bottom panel) demonstrating successful standardization of the corn spectra using the NIST glass standards by the here proposed technique.

Figure 4 shows the same 2 LV model for the prediction of moisture at $\gamma = 10^6$. Both, the reconstructed calibration standard spectra (top left) as well as the reconstructed spectra from the calibration (m5) and the projected validation

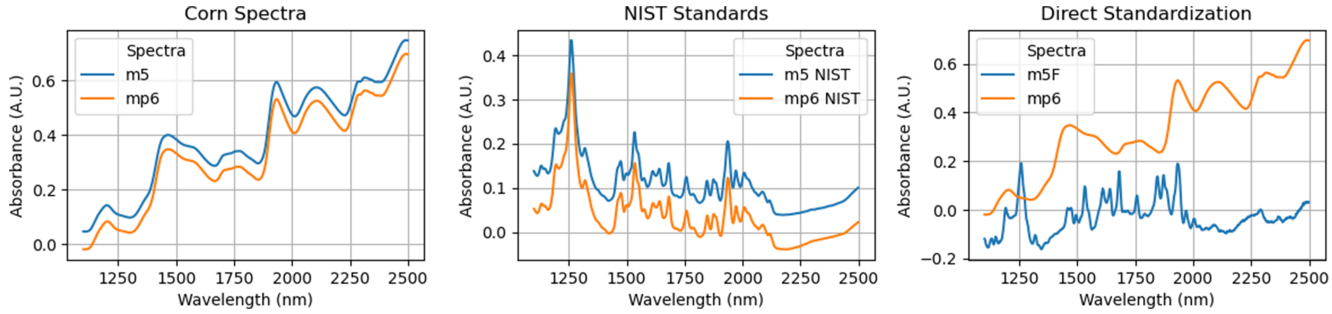

FIGURE 2 Direct standardization. Left: calibration and validation sets corresponding to instruments m5 (blue) and mp6 (orange), respectively. Middle: calibration standards (NIST glass samples) measured on m5 and mp6. Right: m5 spectra after application of direct standardization with $\mathbf{F} = \mathbf{X}^{\dagger}_{\text{m5}_{\text{NIST}}}\mathbf{X}_{\text{mp6}_{\text{NIST}}}$ (m5F)

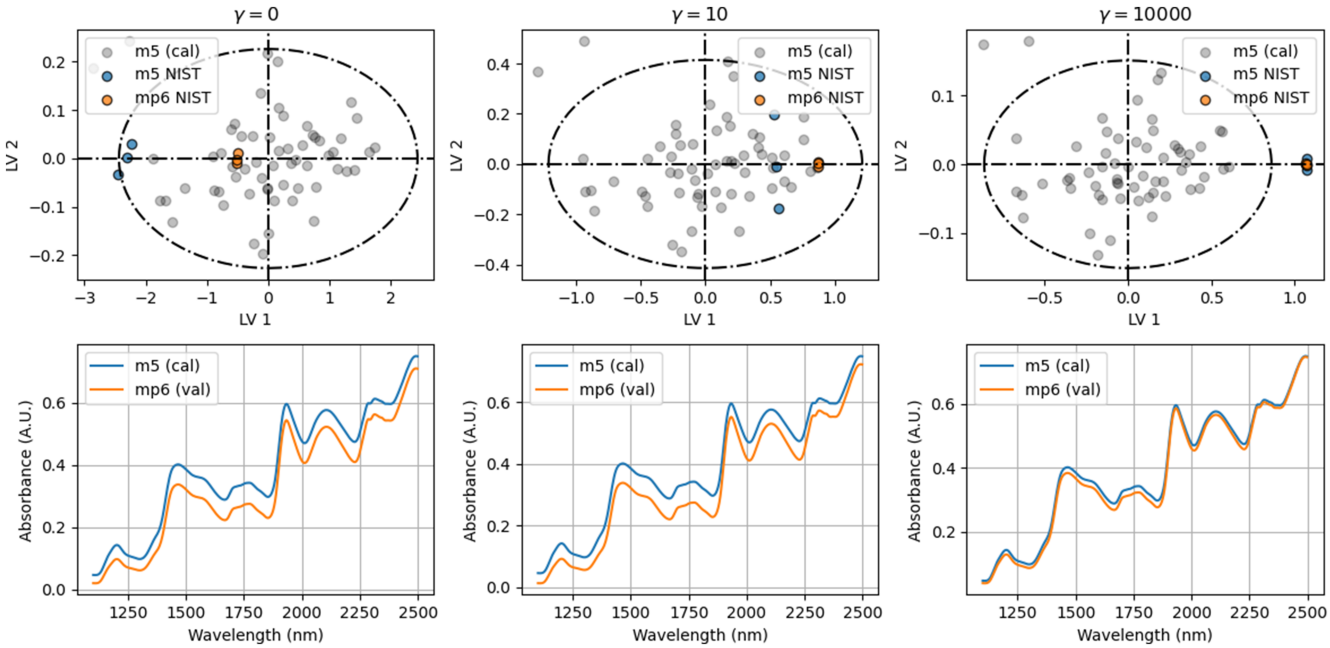

FIGURE 3 Graph-based calibration transfer. Top panel: projections of m5 calibration samples (grey), m5 NIST and mp6 NIST calibration standards with increasing amount of regularization. Bottom panel: corresponding reconstruction of the mean spectra of m5 calibration and mp6 validation samples from a 2 LV model for the prediction of moisture established using the m5 calibration samples

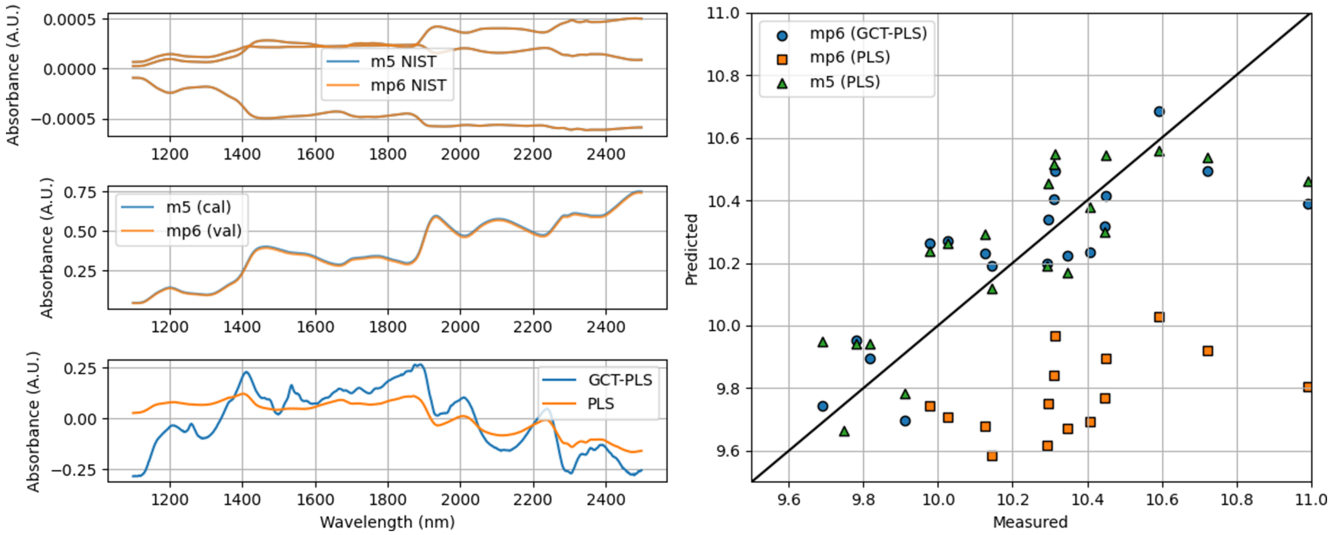

FIGURE 4 Prediction of moisture. Top left: reconstructed calibration standard spectra. Note that primary and secondary calibration standard spectra have been (locally) mean centered for better visibility of the reconstructed spectra. Middle left: reconstructed (mean) corn spectra of the calibration set (m5) and the projected validation samples (mp6). Bottom left: regression coefficients of a 2 latent variable (LV) graph-based calibration transfer-partial least squares (GCT-PLS) ($\gamma = 10^6$) and corresponding PLS model. Right: measured versus predicted moisture contents. Predictions of the GCT-PLS and PLS models on the mp6 validation samples (blue and orange), respectively, and predictions of the PLS model applied to m5 validation samples (green) are shown

samples (mp6) are well aligned and the predictions on the mp6 validation samples thus in good accordance with the respective reference measurements (right plot). The regression vector of the GCT-PLS model displays mostly the same features when compared with the unregularized model but has overall higher variance that can be attributed to the alignment of the matched calibration standards in the LV space. Comparing the accuracy on the validation samples, we have RMSEPs of 0.21, 0.61, and 0.20 for the GCT-PLS, PLS (applied to the mp6 validation samples) and PLS model (applied to the m5 validation samples), respectively, indicating that the transfer model exhibits similar accuracy to the primary (baseline) model applied to the primary validation samples.

4.2 | Number of LVs

Whether CT given one type of samples and another type of calibration standards will succeed in practice will largely depend on the analytical accuracy that is aimed at and, in turn, the number of LVs required to attain that accuracy. Figure 5 shows the reconstructed spectra of the NIST calibration standards as well as the m5 calibration and mp6 validation samples when increasing the number of LVs included in the model. For one and two LVs, setting $\gamma = 10^6$ aligns either the (reconstructed) transfer standards' as well as the corn samples' spectra. In contrast, alignment of the standards does not translate into proper standardization of the corn spectra for a 3 LV model—even at stronger regularization (rightmost plots). This finding indicates that beyond 2 LVs, there is no structure left in the residuals of the calibration standards' spectra (i.e., in the deflated matrices $\boldsymbol{X}_p$ and $\boldsymbol{X}_s$) that encodes differences in the instrumental response of primary and secondary instruments.

In order to verify this hypothesis, we investigated the residuals of $\boldsymbol{X}_s$ and $\boldsymbol{X}_p$ after fitting the first and second LV to the m5 calibration samples using GCT-PLS with $\gamma = 10^6$ (Figure 6). As expected, there is no further structure left in the residuals of the NIST standards' spectra after the second round of deflation that is useful to correct for the differences between the primary and secondary instrument's response. This result indicates that the calibration standards' residuals

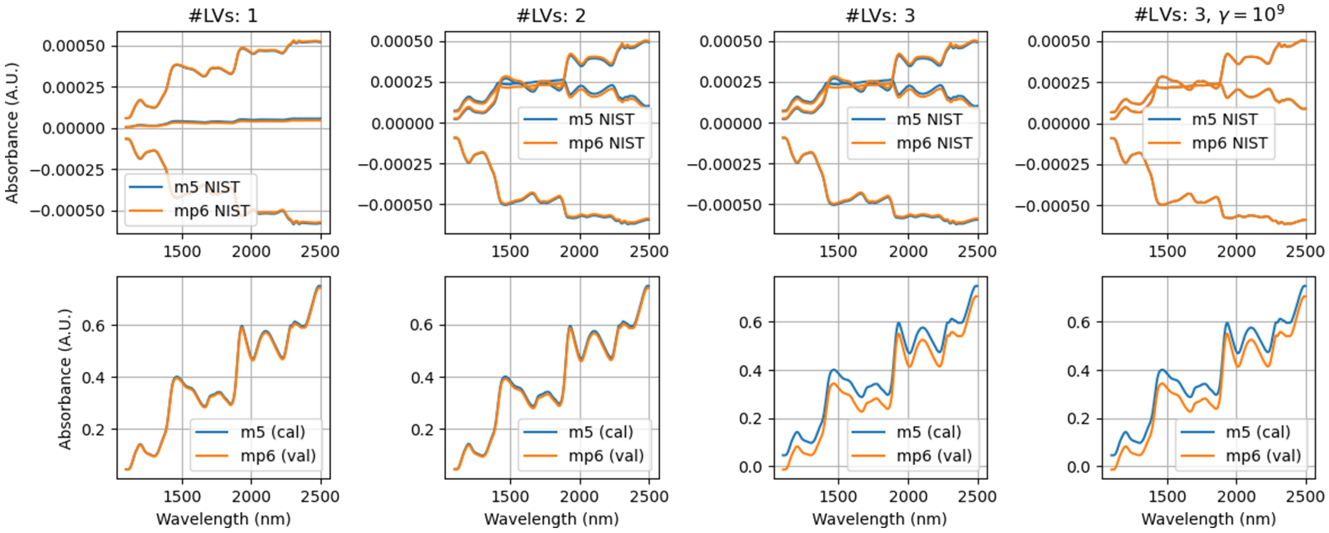

FIGURE 5 Effect of the number of latent variables (LVs) in graph-based calibration transfer (GCT). Reconstruced spectra of the calibration standards (top panel) and corn samples (bottom panel) at increasing number of LVs when modeling moisture. Note that primary and secondary calibration standard spectra have been (locally) mean centered for better visibility of the reconstructed spectra. The number of LVs is indicated column-wise. For the first three columns,$\gamma = 10^6$

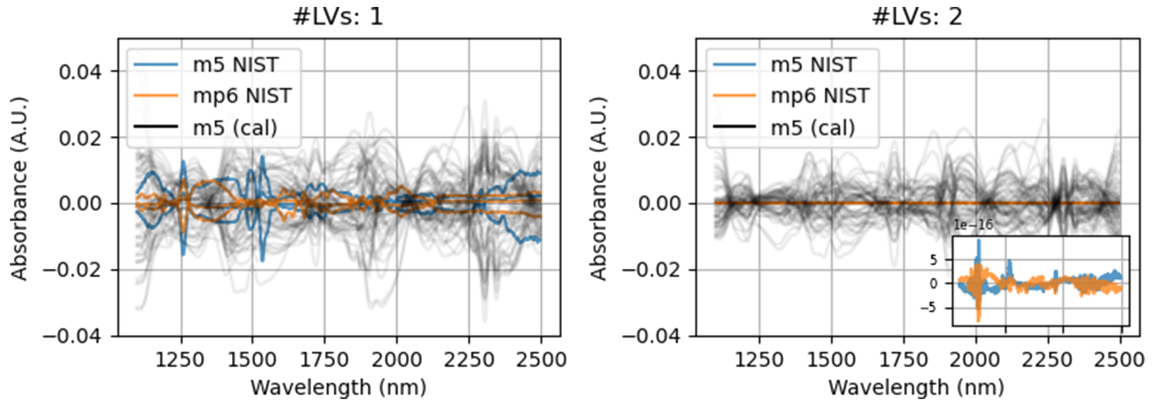

FIGURE 6 Transfer standard residuals. The residuals of m5 NIST and mp6 NIST calibration standards and m5 calibration samples m5 (cal) are shown after fitting the first (left) and second (right) latent variable (LV) using graph-based calibration transfer-partial least squares (GCT-PLS) ($\gamma = 10^6$) for the prediction of moisture. The calibration standards' residuals are magnified in the inset of the right plot for better visibility

represent a useful diagnostic to estimate the number of LVs up to which instrument standardization is feasible given the calibration set (from the primary instrument) and the calibration standards' spectra. For this particular application with corn samples and NIST standards, CT beyond 2 LVs was not feasible for neither of the the responses (moisture, oil, protein, and starch).

4.3 | CT via NIST standards and corn samples

In Figure 7, we compare the RMSEP of GCT-PLS (using the NIST glass standards as CT samples) on the validation data from the secondary instrument with standard PLS (2 LVs). Except for CT between instruments mp5 and mp6, where the difference in the instrumental response is in general small, the accuracy of GCT-PLS models is significantly higher compared with the baseline PLS models and comparable with the accuracy when the secondary PLS model is applied to the secondary validation samples. These findings demonstrate successful application of GCT-PLS for CT using the NIST glass standards for up to 2 LVs.

Interestingly, it was found that the coefficient of determination (R^2) was higher, and the corresponding standard error of calibration (SEC) was lower for GCT-PLS models (compared with the baseline PLS models) in a significant portion of the investigated CT scenarios (Table 1). A possible explanation could be that the proposed regularization tends to shrink the variance of the predictors (i.e., the scores), which eventually leads to overfitting.

Finally, we also investigated application of GCT-PLS when corn samples instead of the NIST glass samples are used as calibration standards (Figure 8 and Table 2). To this end, we employed the Kennard–Stone algorithm in order to select 10 corn spectra from the calibration sets and derive (P)DS standardizations of the (secondary) validation samples and GCT-PLS models. Except for CT from instrument mp5/mp6 to m5 for the prediction of moisture, we found similar accuracy of the theoretically ideal models (i.e., secondary PLS models predicting secondary validation samples) and GCT-PLS models established on the primary calibration samples and the corn standards. In contrast, either DS and PDS performed poorly when transferring models for the prediction of protein and starch (e.g., m5 to mp5), and both were outperformed by GCT-PLS models for the prediction of oil in most instances. Most importantly, these results underpin the weakness of the "traditional" CT approaches, namely, that explicit instrument standardization does not

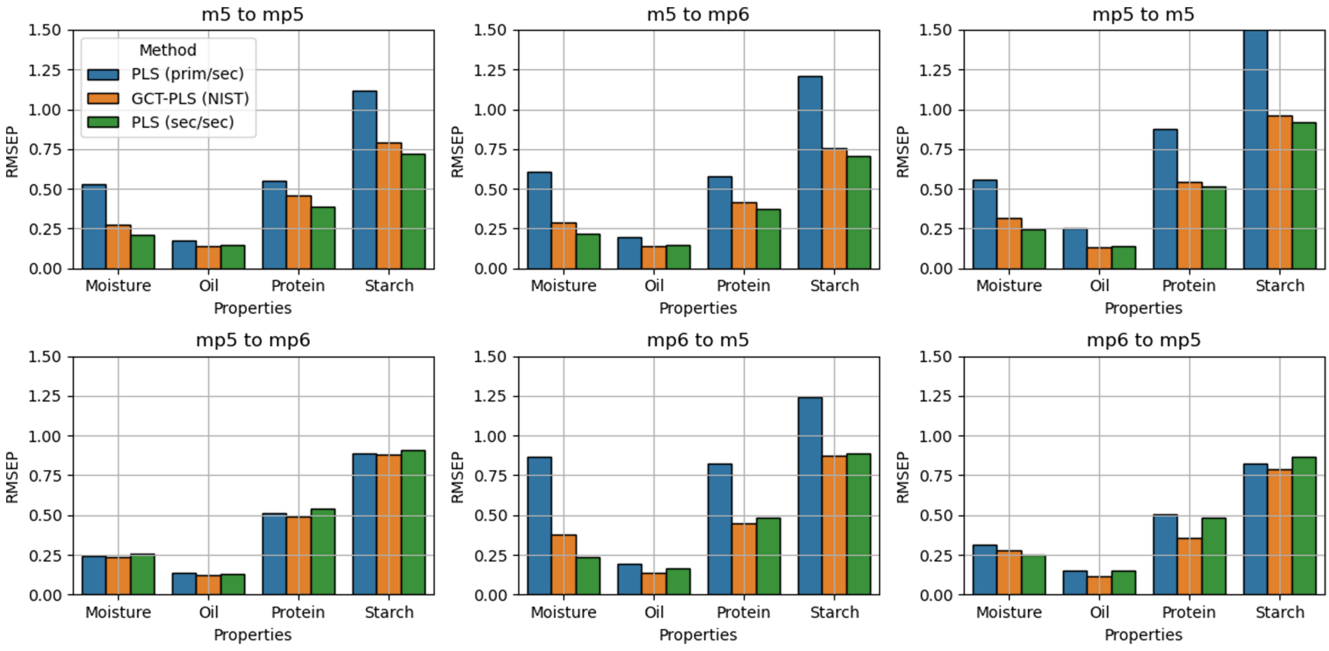

FIGURE 7 Calibration transfer with NIST standards. The root mean squared errors of prediction (RMSEP) are shown for primary partial least squares (PLS) models applied to secondary validation samples (blue), graph-based calibration transfer (GCT)-PLS models employing the NIST standards applied to secondary validation samples (orange) and secondary PLS models applied to secondary validation samples (green). All experiments were conducted with 2 LVs and $\gamma = 10^6$ for GCT-PLS models. The subplot titles indicate CT between primary (e.g., m5) and secondary (mp5) instrument, and the corresponding response values are indicated on the x-axis

TABLE 1 Summary statistics for calibration transfer experiments with NIST standards

		Moisture			Oil			Protein			Starch		
Scenario	**Method**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**
m5 to mp5	PLS (prim/s)	0.75	0.26	0.53	0.36	0.16	0.18	0.44	0.46	0.55	0.34	0.79	1.11
	PLS (s/s)	0.74	0.26	0.21	0.43	0.16	0.14	0.57	0.46	0.38	0.42	0.80	0.72
	GCT-PLS	0.79	0.24	0.26	0.46	0.16	0.14	0.46	0.42	0.43	0.31	0.76	0.80
m5 to mp6	PLS (prim/s)	0.75	0.25	0.61	0.36	0.16	0.20	0.44	0.46	0.58	0.34	0.79	1.20
	PLS (s/s)	0.76	0.25	0.21	0.42	0.15	0.15	0.48	0.45	0.37	0.56	0.70	0.70
	GCT-PLS	0.80	0.23	0.22	0.48	0.15	0.14	0.61	0.41	0.36	0.46	0.74	0.70
mp5 to m5	PLS (prim/s)	0.77	0.25	0.56	0.46	0.16	0.26	0.47	0.42	0.87	0.31	0.74	1.78
	PLS (s/s)	0.78	0.25	0.24	0.40	0.16	0.14	0.42	0.43	0.52	0.34	0.73	0.92
	GCT-PLS	0.81	0.23	0.30	0.52	0.15	0.13	0.57	0.39	0.55	0.44	0.69	1.10
mp5 to mp6	PLS (prim/s)	0.77	0.25	0.24	0.45	0.16	0.13	0.47	0.42	0.51	0.31	0.74	0.88
	PLS (s/s)	0.79	0.24	0.26	0.45	0.16	0.13	0.47	0.41	0.54	0.38	0.72	0.91
	GCT-PLS	0.76	0.26	0.24	0.45	0.16	0.14	0.47	0.42	0.51	0.38	0.72	0.88
mp6 to m5	PLS (prim/s)	0.78	0.24	0.87	0.47	0.15	0.20	0.46	0.43	0.82	0.47	0.70	1.24
	PLS (s/s)	0.77	0.25	0.23	0.43	0.15	0.16	0.42	0.43	0.49	0.33	0.75	0.88
	GCT-PLS	0.82	0.22	0.25	0.55	0.14	0.13	0.66	0.36	0.36	0.56	0.65	0.91
mp6 to mp5	PLS (prim/s)	0.78	0.24	0.31	0.46	0.15	0.15	0.46	0.43	0.50	0.46	0.70	0.82
	PLS (s/s)	0.76	0.25	0.25	0.46	0.15	0.15	0.45	0.43	0.48	0.31	0.75	0.87
	GCT-PLS	0.76	0.25	0.31	0.46	0.15	0.15	0.46	0.42	0.50	0.71	0.56	1.01

Note: Calibration R^2 values, the standard error of calibration (SEC), and the root mean squared error of prediction (RMSEP) are indicated for each transfer scenario and response type partial least squares (PLS, prim/s), PLS (s/s), and graph-based calibration transfer (GCT)-PLS indicate standard primary PLS model evaluated on secondary validation samples, secondary PLS model evaluated on secondary validation samples, and GCT model evaluated on secondary validation samples, respectively.

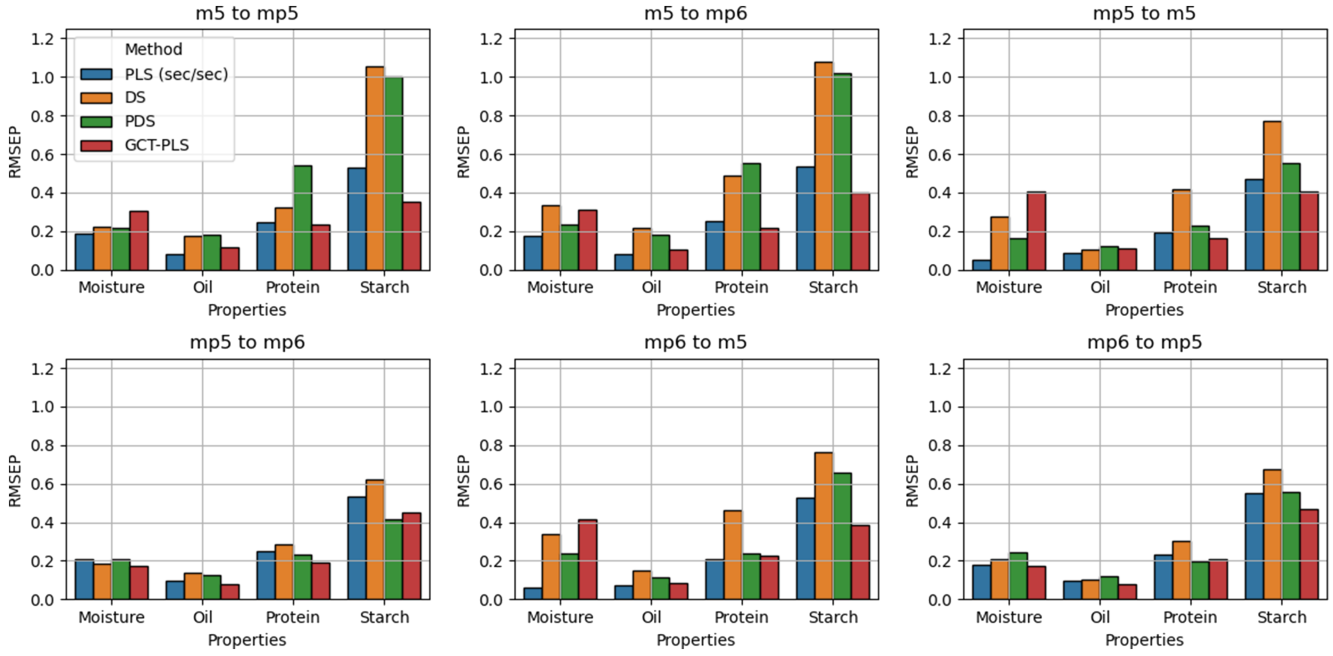

FIGURE 8 Calibration transfer with corn standards. Root mean squared errors of prediction (RMSEP) on the validation samples measured on the secondary instrument are shown for secondary partial least squares (PLS) models applied to secondary validation samples (blue), PLS models established on primary samples standardized with direct standardization (DS) (orange) and piecewise direct standardization (PDS) (green), and graph-based calibration transfer (GCT)-PLS models (red). Five latent variables (LVs) were used for all models, and 10 corn samples, selected from the calibration sets by means of the Kennard–Stone algorithm, were used for standardization

TABLE 2 Summary statistics for calibration transfer experiments with corn standards

		Moisture			Oil			Protein			Starch		
Scenario	**Method**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**	R^2	**SEC**	**RMSEP**
m5 to mp5	PLS	0.92	0.15	0.18	0.84	0.09	0.08	0.93	0.18	0.24	0.85	0.44	0.53
	DS	0.94	0.13	0.22	0.78	0.10	0.17	0.92	0.19	0.32	0.86	0.42	1.05
	PDS	0.99	0.05	0.22	0.91	0.07	0.18	0.94	0.16	0.54	0.88	0.39	1.00
	GCT-PLS	0.96	0.10	0.30	0.90	0.07	0.11	0.97	0.12	0.23	0.93	0.30	0.35
m5 to mp6	PLS	0.90	0.16	0.17	0.82	0.09	0.08	0.93	0.18	0.25	0.86	0.42	0.53
	DS	0.97	0.08	0.33	0.88	0.08	0.21	0.90	0.21	0.48	0.87	0.40	1.07
	PDS	0.99	0.05	0.23	0.91	0.07	0.18	0.94	0.16	0.55	0.88	0.39	1.02
	GCT-PLS	0.96	0.09	0.31	0.91	0.07	0.10	0.97	0.12	0.21	0.93	0.30	0.39
mp5 to m5	PLS	0.99	0.05	0.05	0.91	0.07	0.08	0.93	0.17	0.19	0.86	0.39	0.47
	DS	0.92	0.16	0.27	0.79	0.10	0.10	0.86	0.23	0.41	0.72	0.54	0.77
	PDS	0.93	0.14	0.16	0.87	0.09	0.12	0.92	0.17	0.23	0.84	0.42	0.55
	GCT-PLS	0.94	0.14	0.40	0.88	0.09	0.10	0.95	0.14	0.16	0.90	0.34	0.40
mp5 to mp6	PLS	0.91	0.17	0.20	0.84	0.10	0.09	0.91	0.19	0.24	0.83	0.43	0.53
	DS	0.93	0.15	0.18	0.86	0.09	0.13	0.90	0.21	0.28	0.77	0.49	0.62
	PDS	0.94	0.14	0.21	0.87	0.09	0.12	0.92	0.18	0.23	0.84	0.42	0.41
	GCT-PLS	0.93	0.14	0.17	0.89	0.09	0.08	0.94	0.16	0.18	0.85	0.40	0.45
mp6 to m5	PLS	0.99	0.05	0.06	0.90	0.08	0.07	0.93	0.17	0.20	0.87	0.39	0.52
	DS	0.91	0.16	0.33	0.83	0.09	0.15	0.83	0.26	0.46	0.77	0.50	0.76
	PDS	0.90	0.16	0.23	0.82	0.09	0.11	0.92	0.18	0.24	0.85	0.41	0.65
	GCT-PLS	0.94	0.13	0.42	0.85	0.09	0.08	0.97	0.12	0.22	0.91	0.32	0.38
mp6 to mp5	PLS	0.92	0.15	0.18	0.86	0.09	0.09	0.92	0.18	0.23	0.84	0.43	0.55
	DS	0.90	0.17	0.20	0.81	0.10	0.10	0.86	0.24	0.30	0.78	0.50	0.67
	PDS	0.90	0.16	0.24	0.82	0.10	0.11	0.92	0.18	0.19	0.85	0.41	0.55
	GCT-PLS	0.92	0.15	0.17	0.88	0.09	0.07	0.95	0.14	0.20	0.88	0.38	0.47

Note: Calibration R^2 values, the standard error of calibration (SEC), and the root mean squared error of prediction (RMSEP) are indicated for each transfer scenario and response type. Partial least squares (PLS), direct standardization (DS), piecewise direct standardization (PDS), and graph-based calibration transfer (GCT)-PLS indicate secondary PLS models evaluated on secondary validation samples, direct, and piecewise direct standardization and GCT-PLS models evaluated on secondary validation samples, respectively. All models were fitted with 5 LVs.

take into account the corresponding calibration model. As a consequence, a (P)DS standardization might work well for one response (e.g., moisture) but fails on another one (e.g., protein).

5 | CONCLUSION

We have here introduced a graph-based approach to address the CT problem in analytical chemistry, which, unlike most other SoA methods, implicitly standardizes the instrumental responses of primary and secondary instruments while modeling the property of interest. The main benefit of our approach lies in the fact that calibration standards and calibration samples must not share the same spectral features, because the instrumental differences are implicitly modeled in the space of the calibration samples. This was illustrated by employing the NIST glass standards for transferring calibrations for the prediction of corn properties between primary and secondary instruments. We have further shown experimentally that the residuals of the calibration standard spectra in the GCT model indicate how many LVs the chosen calibration standards permit to transfer from one instrument to the other, which is an important quantity when assessing the feasibility of CT under given accuracy requirements. Finally, we have shown that GCT-PLS also compares favorably with standard CT techniques like (P)DS when calibration and CT samples are of the same type.

Future work should further investigate the capability of GCT-PLS to transfer calibrations given one type of calibration samples and another type of CT standards in order to better understand under which conditions models can be successfully transferred in this setting. Finally, ongoing work also considers the frequently encountered situation of wavelength registry shifts between instruments.[27]

ACKNOWLEDGEMENT

The research reported in this work has been partly funded by BMVIT, BMDW, and the Province of Upper Austria in the frame of the COMET Program managed by the FFG. The COMET Centre CHASE is funded within the framework of COMET—Competence Centers for Excellent Technologies by BMVIT, BMDW, and the Federal Provinces of Upper Austria and Vienna. The COMET program is run by FFG.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/cem.3319.

ORCID

Ramin Nikzad-Langerodi https://orcid.org/0000-0003-3495-8949

REFERENCES

1. Burns D, Ciurczak E. *Handbook of Near-Infrared Analysis*. Boca Raton, Florida, USA: CRC Press; 2007.
2. Feudale RN, Woody NA, Tan H, Myles AJ, Brown SD, Ferré J. Transfer of multivariate calibration models: a review. *Chemom Intell Lab Syst*. 2002;64(2):181-192. http://www.sciencedirect.com/science/article/pii/S0169743902000850
3. Workman JJ. A review of calibration transfer practices and instrument differences in spectroscopy. *Appl Spectrosc*. 2018;72(3):340-365.
4. Bouveresse E, Hartmann C, Massart DL, Last IR, Prebble KA. Standardization of near-infrared spectrometric instruments. *Anal Chem*. 1996;68(6):982-990. https://doi.org/10.1021/ac9510595
5. Nikzad-Langerodi R, Lughofer E, Cernuda C, Reischer T, Kantner W, Pawliczek M, Brandstetter M. Calibration model maintenance in melamine resin production: integrating drift detection, smart sample selection and model adaptation. *Anal Chimica Acta*. 2018;1013:1-12. (featured article).
6. Kalivas JH, Siano GG, Andries E, Goicoechea HC. Calibration maintenance and transfer using tikhonov regularization approaches. *Appl Spectrosc*. 2009;63(7):800-809. https://doi.org/10.1366/000370209788701206, PMID: 19589218.
7. Andrew A, Fearn T. Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation. *Chemom Intell Lab Syst*. 2004;72(1):51-56. http://www.sciencedirect.com/science/article/pii/S0169743904000516
8. Malli B, Birlutiu A, Natschläger T. Standard-free calibration transfer—an evaluation of different techniques. *Chemom Intell Lab Syst*. 2017;161:49-60. http://www.sciencedirect.com/science/article/pii/S0169743916305494
9. Poerio DV, Brown SD. Dual-domain calibration transfer using orthogonal projection. *Appl Spectrosc*. 2018;72(3):378-391. PMID: 28707982.
10. Andries E. Penalized eigendecompositions: motivations from domain adaptation for calibration transfer. *J Chemometr*. 2017;31(4):e2818.
11. Nikzad-Langerodi R, Zellinger W, Lughofer E, Saminger-Platz S. Domain-invariant partial-least-squares regression. *Anal Chem*. 2018;90(11):6693-6701.
12. Nikzad-Langerodi R, Zellinger W, Saminger-Platz S, Moser B. Domain-invariant regression under Beer-Lambert's law. In: 2019 18th IEEE International Conference on Machine Learning and Applications (icmla); 2019:581-586.
13. Wang Y, Veltkamp DJ, Kowalski BR. Multivariate instrument standardization. *Anal Chem*. 1991;63(23):2750-2756. https://doi.org/10.1021/ac00023a016
14. Wise BM, Martens H, Høy M, Bro R, Brockhoff PB. Calibration transfer by generalized least squares. In: Proceedings of the Seventh Scandinavian Symposium on chemometrics (SSC7); 2001; Copenhagen, Denmark:19-23.
15. Martens H, Høy M, Wise BM, Bro R, Brockhoff PB. Pre-whitening of data by covariance-weighted pre-processing. *J Chemometr*. 2003;17(3):153-165. https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.780
16. Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometr Intell Lab Syst*. 1998;44(1):229-244. http://www.sciencedirect.com/science/article/pii/S0169743998001129
17. Igne B, Roger J-M, Roussel S, Bellon-Maurel V, Hurburgh CR. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemometr Intell Lab Syst*. 2009;99(1):57-65. http://www.sciencedirect.com/science/article/pii/S016974390900149X
18. Du W, Chen Z-P, Zhong L-J, et al. Maintaining the predictive abilities of multivariate calibration models by spectral space transformation. *Anal Chimica Acta*. 2011;690(1):64-70. http://www.sciencedirect.com/science/article/pii/S0003267011001863
19. Peng J, Peng S, Jiang A, Tan J. Near-infrared calibration transfer based on spectral regression. *Spectrochim Acta A Mol Biomol Spectrosc*. 2011;78(4):1315-1320. http://www.sciencedirect.com/science/article/pii/S1386142511000072
20. Zhao Y, Yu J, Shan P, Zhao Z, Jiang X, Gao S. Pls subspace-based calibration transfer for near-infrared spectroscopy quantitative analysis. *Molecules (Basel, Switzerland)*. 2019;24(7):1289. https://pubmed.ncbi.nlm.nih.gov/30987017

21. Skotare T, Nilsson D, Xiong S, Geladi P, Trygg J. Joint and unique multiblock analysis for integration and calibration transfer of nir instruments. *Anal Chem*. 2019;91(5):3516-3524. https://doi.org/10.1021/acs.analchem.8b05188
22. Folch-Fortuny A, Vitale R, de Noord OE, Ferrer A. Calibration transfer between nir spectrometers: new proposals and a comparative study. *J Chemometr*. 2017;31(3):e2874. https://doi.org/10.1002/cem.2874
23. He X, Niyogi P. Locality preserving projections. In: Proceedings of the 16th International Conference on Neural Information Processing Systems, NIPS'03. MIT Press; 2003; Cambridge, MA, USA:153-160.
24. Zhong B, Wang J, Zhou J, Wu H, Jin Q. Quality-related statistical process monitoring method based on global and local partial least-squares projection. *Industr Eng Chem Res*. 2016;55(6):1609-1622. https://doi.org/10.1021/acs.iecr.5b02559
25. Lughofer E, Nikzad-Langerodi R. Robust generalized fuzzy systems training from high-dimensional time-series data using local structure preserving pls. *IEEE Trans Fuzzy Syst*. 2019;28:1-1.
26. Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics*. 1969;11(1):137-148. https://www.tandfonline.com/doi/abs/10.1080/00401706.1969.10490666
27. Isaksson T, Yang H, Kemeny GJ, Jackson RS, Wang Q, Alam MK, Griffiths PR. Accurate wavelength measurements of a putative standard for near-infrared diffuse reflection spectrometry. *Appl Spectr*. 2003;57(2):176-185.

How to cite this article: Nikzad-Langerodi R, Sobieczky F. Graph-based calibration transfer. *Journal of Chemometrics*. 2020;e3319. https://doi.org/10.1002/cem.3319

APPENDIX: A

The calculation of the optimal weights vector w^* is easily obtained from the necessary condition

$$\nabla_w\left(||\mathbf{X}-\mathbf{y}\mathbf{w}^T||_F^2+\gamma J(w)\right)=0. \tag{A1}$$

Due to $\mathrm{Tr}[\mathbf{A}\mathbf{A}^T]=||\mathbf{A}||_F^2$, and the cyclicity of the trace ($\mathrm{Tr}[\mathbf{x}\mathbf{y}^T]=\mathbf{y}^T\mathbf{x}$), the left-hand side can be written as

$$\begin{aligned}\nabla_w\left(\mathrm{Tr}\left[\mathbf{X}\mathbf{X}^T-\mathbf{X}\mathbf{w}\mathbf{y}^T-\mathbf{y}\mathbf{w}^T\mathbf{X}^T+\mathbf{y}\mathbf{w}^T\mathbf{w}\mathbf{y}^T\right]+\gamma\mathbf{w}^T\boldsymbol{\Gamma}\mathbf{w}\right) &= 0-\nabla_w(2\mathrm{Tr}[\mathbf{X}\mathbf{w}\mathbf{y}^T]+\mathbf{y}^T\mathbf{y}\mathbf{w}^T\mathbf{w}+\gamma\mathbf{w}^T\boldsymbol{\Gamma}\mathbf{w})\\ =\nabla_w(-2\mathbf{y}^T\mathbf{X}\mathbf{w}+\mathbf{y}^T\mathbf{y}\mathbf{w}^T\mathbf{w}+\gamma\mathbf{w}^T\boldsymbol{\Gamma}\mathbf{w}) &= 2\left(-\mathbf{X}^T\mathbf{y}+\mathbf{y}^T\mathbf{y}\mathbf{w}+\gamma\boldsymbol{\Gamma}\mathbf{w}\right).\end{aligned}$$

Solving (A1) for $\mathbf{w}$ yields $\mathbf{w}=\left(\mathbf{I}+\frac{\gamma}{\mathbf{y}^T\mathbf{y}}\boldsymbol{\Gamma}\right)^{-1}\mathbf{X}^T\mathbf{y}$, which is the column-vector equivalent to (2).