Received: 17 May 2018 | Revised: 2 August 2018 | Accepted: 15 August 2018

DOI: 10.1002/cem.3080

RESEARCH ARTICLE

WILEY Journal of CHEMOMETRICS

Sample and feature augmentation strategies for calibration updating

Erik Andries[1] | John H. Kalivas[2] | Anit Gurung[2]

[1]Department of Mathematics, Science and Engineering, Central New Mexico Community College, Albuquerque, New Mexico

[2]Department of Chemistry, Idaho State University, Pocatello, Idaho

Correspondence
Erik Andries, Department of Mathematics, Science and Engineering, Central New Mexico Community College, 525 Buena Vista Dr SE, Albuquerque, NM 87106, USA.
Email: eandries@cnm.edu

Funding information
National Science Foundation , Grant/Award Number: CHE-1506417

Handling Editor: Abel Marie-Laure

Calibration updating—transfer and/or maintenance—has historically been implemented using a simple but effective technique: in addition to primary samples, include a small number of secondary samples and weight them. It would be beneficial if these classical weighting techniques could be enhanced. Moreover, it would be ideal if we could only use secondary spectra without reference measurements. In this paper, we examine multiple calibration updating scenarios involving unlabeled and labeled secondary spectra. First, we propose three new updating approaches involving sample augmentation whereby unlabeled secondary spectra are used to construct an "undesirable" subspace. This subspace is then used to steer the model vector away from a spectroscopically undesirable solution. Second, we propose two new feature augmentation approaches using labeled secondary samples. These three approaches involves the sum of two model vectors, a dedicated primary model vector plus a perturbation vector, that can accommodate new secondary samples. We rigorously vet the proposed approaches across two near-infrared (NIR) data sets and across multiple data splits. Out of all of the approaches examined, one feature augmentation approach provides improved results compared with existing approaches, and one sample augmentation approach utilizing only unlabeled secondary spectra appears promising.

KEYWORDS

calibration updating, domain adaptation, feature augmentation, sample augmentation, unlabeled data

1 | INTRODUCTION

The normal process of building a spectral calibration model typically requires a large set of samples such that all variances are included in future predictions. However, once such a primary calibration model has been built, circumstances can cause the model to become invalid. For example, instrumental drift or uncalibrated spectral features appearing in new secondary samples can occur later in time. Alternatively, an unknown secondary sample could be measured on an instrument other than the instrument the calibration model was built on. Another situation invalidating a primary model are seasonal changes for agricultural analysis. For these and other situations, the instrument must be recalibrated to accommodate new conditions. In theory, the remedy would be to include a large number of new calibration samples along with their reference measurements. In practice, though, the inclusion of such samples is often costly and lengthy in terms of laboratory time. Hence, mechanisms are needed to update the current model to include the new chemical, physical,

environmental, and/or instrumental effects not spanning the current primary calibration domain. We use calibration updating as an umbrella term for adapting secondary to primary situations.

This paper has two objectives: (*a*) combine data from two sources—a large set of primary samples and a smaller set of secondary samples into a single linear regression framework—and (*b*) make predictions for new secondary samples. The number of secondary samples with reference measurements in the secondary calibration set needs to be small in number relative to the number of primary samples for the following reasons. In many real world applications, it is often difficult or expensive to acquire additional reference measurements. In some cases, acquiring additional reference measurements is not an issue but not desirable, say, from a marketing perspective or from the actual use-case scenario of a device. Instead of using secondary samples with reference values, this paper also studies methods and situations where the secondary reference values are not known, ie, unlabeled samples.

The differences between primary and secondary samples follow a continuum from similar to dissimilar. By similar, we mean that the primary and secondary samples are drawn from the same statistical distribution and share the same sources of variation. If the primary and secondary samples are highly similar, then one can simply pool all of the samples together. If the secondary samples are radically dissimilar from the primary samples, then one can ignore the primary samples and build a calibration model solely on the secondary samples … provided there are enough samples. Many settings in chemometrics are intermediate between these two extremes: although we expect the secondary samples to be dissimilar (but not that dissimilar) to the primary samples, the primary samples should still provide leverage such that improved prediction can be obtained for secondary samples. Although the primary and secondary samples often share considerable spectral overlap across wavelength bands and are highly correlated, the subtle differences are enough to wreck havoc on a prediction model based solely on primary samples.

Calibration updating has a long history and is well studied in chemometrics and spectroscopy; see other studies[1-18] and references therein. We narrow the scope of inquiry to sample and feature augmentations, a class of calibration updating techniques that vertically and/or horizontally concatenate secondary samples to primary samples. As will be illustrated later on in this paper, this concatenation of secondary samples results in a penalized least squares problem. Note that we are not investigating the following calibration updating approaches:

- Approaches that require a standardization set.[1-5,8] A standardization set is a common set of samples measured across two or more instruments (or samples measured across two or more different measuring conditions). Since each standardization sample has the same reference value across instruments, the variability of the spectral measurements should largely reflect instrument-to-instrument difference. To establish transfer parameters, the standardization samples need to be representative of the entire experimental regime and stable enough over time between situations in which the standardization is performed. For many situations, the creation of a standardization set is not possible.
- Spectral preprocessing techniques.[7,9,19-22] These techniques refer to approaches (eg, wavelets) that transform the spectra to minimize domain differences between the primary and secondary samples without the use of a standardization set.

This is not to say that one cannot utilize standardization sets or preprocessing techniques in tandem with sample and feature augmentations, but those investigations are outside the scope of this paper.

The term model, as used here in this paper, refers to a linear model, ie, a system of linear equations coupling together both primary and secondary samples in a particular fashion. What is of importance is how one constructs the linear system such that a priori chemical and/or spectroscopic information can be meaningfully incorporated. Of less importance is the particular regression algorithm that is subsequently applied to the linear system. We are agnostic with respect to regression algorithms in part because each discipline has its own set of "go-to" methods to apply. In short, the art and craft of calibration updating resides in the construction of a linear model. For the calibration updating models discussed in later sections, partial least squares (PLS) will be the default regression algorithm applied to the linear model.

This paper is organized as follows. Section 2 describes the four baseline models that the proposed updating methods will be compared against. Section 3 describes a class of calibration updating models with unlabeled data collectively referred to as sample augmentation models. Three methods will be described here. Section 4 describes feature augmentation models in which the number of features (wavelengths) are increased by horizontally concatenating secondary spectra (or matrices derived from secondary spectra) to primary spectra and reference measurements. Two methods are described. Section 5 introduces the data sets to be examined and how they will be partitioned for performance assessment. Section 6 discusses performance results across the various calibration updating models. Section 7 concludes the paper and discusses future work.

TABLE 1 Summary of updating methods and associated equations[a]

Updating method	Section number	Calibration model	Prediction
Baseline methods			
PPS	2.1	$\mathbf{X}_\mathrm{P}\mathbf{b} = \mathbf{y}_\mathrm{P}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{P}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{P}^\mathrm{y}$
SPS, SSPS	2.2, 2.3	$\mathbf{X}_\mathrm{S}\mathbf{b} = \mathbf{y}_\mathrm{S}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{S}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{S}^\mathrm{y}$
LMC	2.4	$\begin{bmatrix}\mathbf{X}_\mathrm{P}\\ \lambda\mathbf{X}_\mathrm{S}\end{bmatrix}\mathbf{b} = \begin{bmatrix}\mathbf{y}_\mathrm{P}\\ \lambda\mathbf{y}_\mathrm{S}\end{bmatrix}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{S}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{S}^\mathrm{y}$
Sample augmented updating methods			
NAR	3.1, 3.2	$\begin{bmatrix}\mathbf{X}_\mathrm{P}\\ \tau\mathbf{R}\end{bmatrix}\mathbf{b} = \begin{bmatrix}\mathbf{y}_\mathrm{P}\\ \mathbf{0}\end{bmatrix}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{P}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{P}^\mathrm{y}$
Feature augmented updating methods			
FA-1	4.2	$\begin{bmatrix}\mathbf{X}_\mathrm{P} & \mathbf{0}\\ \lambda\mathbf{X}_\mathrm{S} & \lambda\mathbf{X}_\mathrm{S}\end{bmatrix}\begin{bmatrix}\mathbf{b}_\mathrm{P}\\ \mathbf{b}_\mathrm{S}\end{bmatrix} = \begin{bmatrix}\mathbf{y}_\mathrm{P}\\ \lambda\mathbf{y}_\mathrm{S}\end{bmatrix}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{S}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{S}^\mathrm{y}$
FA-2	4.2	$\begin{bmatrix}\mathbf{X}_\mathrm{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}_\mathrm{P}\\ \lambda\mathbf{X}_\mathrm{S} & \lambda\mathbf{X}_\mathrm{S}\end{bmatrix}\begin{bmatrix}\mathbf{b}_\mathrm{P}\\ \mathbf{b}_\mathrm{S}\end{bmatrix} = \begin{bmatrix}\mathbf{y}_\mathrm{P}\\ \mathbf{0}\\ \lambda\mathbf{y}_\mathrm{S}\end{bmatrix}$	$\hat{y}_\mathrm{S} = \left(\mathbf{x}_\mathrm{S} - \boldsymbol{\mu}_\mathrm{S}^\mathrm{X}\right)^\mathrm{T}\mathbf{b} + \mu_\mathrm{S}^\mathrm{y}$

[a]See respective sections for definitions.

We now discuss notation. Symbols that are not boldface represent scalars (x or P). Lowercase and uppercase boldface symbols represent column vectors ($\mathbf{x}$) or matrices ($\mathbf{X}$). All vectors are column vectors unless noted otherwise. The superscripted symbols $^\mathrm{T}$ and $^{-1}$ indicate the transpose and inverse, respectively, of a vector or matrix. The matrices $\mathbf{I}$ and $\mathbf{0}$ and vectors $\mathbf{1}_\mathrm{d}$ and $\mathbf{0}_\mathrm{d}$ indicate the identity matrix, a matrix of zeros, and a column vector of d ones and zeros, respectively. The comma and semicolon indicate the horizontal and vertical concatenation (or stacking) of matrix/vector entries. For example, it will be convenient to represent a $n \times d$ matrix $\mathbf{X}$ of spectra by concatenating column vectors $\mathbf{x}_j = [x_{j1}; x_{j2}; \ldots; x_{jd}]$ (or $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jd}]^\mathrm{T}$) such that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^\mathrm{T}$ whereby $\mathbf{x}_j$ corresponds to the jth spectrum. The vector $\mathbf{y} = [y_1; y_2; \ldots; y_n]$ (or $\mathbf{y} = [y_1, y_2, \ldots, y_n]^\mathrm{T}$) represents the response variable (eg, reference measurements such as analyte concentrations). We only consider the case where each spectrum is associated with an univariate response variable, not multiple response variables. In this paper, $\mathbf{X}_\mathrm{P} = [\mathbf{x}_{(\mathrm{P},1)}, \cdots, \mathbf{x}_{(\mathrm{P},n_\mathrm{P})}]^\mathrm{T}$ and $\mathbf{y}_\mathrm{P} = [y_{(\mathrm{P},1)}, \cdots, y_{(\mathrm{P},n_\mathrm{P})}]^\mathrm{T}$ correspond to n_P samples of primary calibration spectra and their corresponding reference measurements, respectively. Similarly, $\mathbf{X}_\mathrm{S} = [\mathbf{x}_{(\mathrm{S},1)}, \cdots, \mathbf{x}_{(\mathrm{S},n_\mathrm{S})}]^\mathrm{T}$ and $\mathbf{y}_\mathrm{S} = [y_{(\mathrm{S},1)}, \cdots, y_{(\mathrm{S},n_\mathrm{S})}]^\mathrm{T}$ correspond to n_S samples of secondary calibration spectra and their respective reference measurements. We will heretofore refer to samples with and without reference measurements as labeled and unlabeled samples, respectively. Moreover, we will also require the use of n_SU unlabeled secondary samples in the matrix $\mathbf{X}_\mathrm{SU}$. It is assumed that the number of secondary calibration samples is small in size relative to the number of primary calibration samples and the number of unlabeled secondary samples, ie, $n_\mathrm{S} \ll n_\mathrm{P}$ and $n_\mathrm{S} \ll n_\mathrm{SU}$. In addition, we assume the primary and secondary spectra span the same wavelength range at the wavelength intervals. The updating methods discussed in the following sections are tabulated in Table 1.

2 | BASELINE UPDATING MODELS

As mentioned in section 1, PLS will be the regression approach applied to all of the linear systems in Table 1. The number of latent vectors associated with PLS will be denoted by the integer k. In this section, we will introduce four baseline methods by which the other updating methods will be judged: primary predicting secondary (PPS), secondary predicting secondary (SPS), small secondary predicting secondary (SSPS), and local mean centering (LMC).

2.1 | Primary predicting secondary

Primary predicting secondary corresponds to no calibration updating whatsoever—the calibration data consist only of a large pool of primary samples. As a result, we solve the minimization problem (and corresponding linear system) for the model vector $\mathbf{b}$:

$$\min_{\mathbf{b}} ||\mathbf{X}_\mathrm{P}\mathbf{b} - \mathbf{y}_\mathrm{P}||_2^2 \quad \Longleftrightarrow \quad \mathbf{X}_\mathrm{P}\mathbf{b} = \mathbf{y}_\mathrm{P}. \tag{1}$$

The matrix $\mathbf{X}_P$ and vectors $\mathbf{b}$ and $\mathbf{y}_P$ have dimensions $n_P \times d$, $d \times 1$ and $n_P \times 1$, respectively. Here, we assume that both $\mathbf{X}_P$ and $\mathbf{y}_P$ have been mean centered, ie,

$$\begin{aligned} \mathbf{X}_P &:= \mathbf{X}_P - \mathbf{1}_{n_p}(\boldsymbol{\mu}_P^X)^T, \quad \boldsymbol{\mu}_P^X = \frac{1}{n_P}(\mathbf{X}_P)^T\mathbf{1}_{n_p}, \\ \mathbf{y}_P &:= \mathbf{y}_P - \mathbf{1}_{n_p}(\mu_P^y)^T, \quad \mu_P^y = \frac{1}{n_P}(\mathbf{y}_P)^T\mathbf{1}_{n_p}. \end{aligned} \tag{2}$$

(The symbol ":=" is meant to represent *assignment* as in "the left-hand side is assigned the value of the right-hand side.") Prediction on a new secondary sample $\mathbf{x}_S$ proceeds by mean centering the secondary sample using the primary mean: $\hat{y}_S = \left(\mathbf{x}_S - \boldsymbol{\mu}_P^X\right)^T\mathbf{b} + \mu_P^y$. We expect any worthwhile calibration updating approach to outperform PPS.

Note that Equation 1 attempts to find a vector $\mathbf{b} = [b_1, b_2, \ldots, b_n]^T$ that adequately models the primary data with respect to prediction on a novel primary sample $\mathbf{x}_P$. For the model vector $\mathbf{b}$ to give preference to secondary data, a penalty term is often appended to the minimization:

$$||\mathbf{X}_P\mathbf{b} - \mathbf{y}_P||_2^2 + \lambda^2 P, \tag{3}$$

for some suitably chosen P. Different choices for P encode different mechanisms for updating a model to secondary data. Such penalty choices will be examined later in sections 2.4, 3, and 4.

2.2 | Secondary predicting secondary

Secondary predicting secondary corresponds to the best case scenario with respect to prediction on new secondary spectra: the calibration samples consist of a large pool of secondary samples. The ideal calibration updating method seeks to match (but not likely to exceed) the performance of SPS. In terms of a linear system, we solve

$$\min_{\mathbf{b}} ||\mathbf{X}_S\mathbf{b} - \mathbf{y}_S||_2^2 \quad \Longleftrightarrow \quad \mathbf{X}_S\mathbf{b} = \mathbf{y}_S, \tag{4}$$

for the model vector $\mathbf{b}$. The matrix $\mathbf{X}_S$ and vectors $\mathbf{b}$ and $\mathbf{y}_S$ have dimensions $n_S \times d$, $d \times 1$, and $n_S \times 1$, respectively. Similar to PPS, we assume that both $\mathbf{X}_S$ and $\mathbf{y}_S$ have been mean centered, ie,

$$\begin{aligned} \mathbf{X}_S &:= \mathbf{X}_S - \mathbf{1}_{n_s}(\boldsymbol{\mu}_S^X)^T, \quad \boldsymbol{\mu}_S^X = \frac{1}{n_S}(\mathbf{X}_S)^T\mathbf{1}_{n_s}, \\ \mathbf{y}_S &:= \mathbf{y}_S - \mathbf{1}_{n_s}(\mu_S^y)^T, \quad \mu_S^y = \frac{1}{n_S}(\mathbf{y}_S)^T\mathbf{1}_{n_s}. \end{aligned} \tag{5}$$

Prediction on a new secondary sample $\mathbf{x}_S$ proceeds by mean centering the secondary sample using the secondary means: $\hat{y}_S = \left(\mathbf{x}_S - \boldsymbol{\mu}_S^X\right)^T\mathbf{b} + \mu_S^y$. In practice, one never knows the level of performance associated with SPS because one generally never has that many labeled secondary samples to begin with. (If they did, then there would be no need for calibration updating in the first place.) Secondary predicting secondary serves as the theoretical ideal that all calibration updating methods aspire to.

2.3 | Small secondary predicting secondary

Small secondary predicting secondary is the same as SPS except that a much smaller set of secondary samples is used. Here, we ask the question: instead of a full-blown recalibration requiring many secondary reference values, can one achieve similar performance results using the same smaller set of labeled secondary samples used in the sample augmentation method? If so, then there is no need to use an updating method that pools together a large number of labeled primary samples with the same small number of labeled secondary samples.

2.4 | Local mean centering

Here, we add a small set of labeled secondary samples to a large pool of labeled primary samples such that $n_S \ll n_P$. The labeled secondary samples are then weighted by a value denoted by λ. The resulting model vector $\mathbf{b}$ depends upon two tuning parameters: the number of PLS latent vectors k and the weight λ. The use of the word *local* refers to the separate mean centering of the primary and secondary samples:

$$\begin{aligned} \mathbf{X}_P &:= \mathbf{X}_P - \mathbf{1}_{n_p}(\boldsymbol{\mu}_P^X)^T, \qquad \mathbf{y}_P := \mathbf{y}_P - \mathbf{1}_{n_p}(\mu_P^y)^T, \\ \mathbf{X}_S &:= \mathbf{X}_S - \mathbf{1}_{n_s}(\boldsymbol{\mu}_S^X)^T, \qquad \mathbf{y}_S := \mathbf{y}_S - \mathbf{1}_{n_s}(\mu_S^y)^T. \end{aligned} \tag{6}$$

One then solves the following *sample augmented* for the model vector $\mathbf{b}$

$$\min_{\mathbf{b}} \left(||\mathbf{X}_P\mathbf{b} - \mathbf{y}_P||_2^2 + \lambda^2 ||\mathbf{X}_S\mathbf{b} - \mathbf{y}_S||_2^2 \right) \iff \begin{bmatrix} \mathbf{X}_P \\ \lambda\mathbf{X}_S \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y}_P \\ \lambda\mathbf{y}_S \end{bmatrix}, \tag{7}$$

whereby weighted secondary samples are vertically stacked underneath primary samples. The weighting by λ in Equation 7 is performed on mean-centered secondary samples. That is, model updating is largely achieved by first moving the centroids of the primary and secondary spectra (via separate mean centering) independently to the origin and then scaling the size of the secondary samples. The prediction on a new secondary spectrum $\mathbf{x}_S$ uses the secondary means: $\hat{y}_S = \left(\mathbf{x}_S - \boldsymbol{\mu}_S^X\right)^T \mathbf{b} + \mu_S^y$.

When $\lambda > 0$ in Equation 7, the resulting model vector $\mathbf{b}$ obviously depends upon both $\{\mathbf{X}_P, \mathbf{y}_P\}$ and $\{\mathbf{X}_S, \mathbf{y}_S\}$. When $\lambda = 0$, the linear system reduces to $\mathbf{X}_P\mathbf{b} = \mathbf{y}_P$ and the vector $\mathbf{b}$ models only primary samples. However, the prediction $\hat{y}_S = \left(\mathbf{x}_S - \boldsymbol{\mu}_S^X\right)^T \mathbf{b} + \mu_S^y$ still utilizes the means from the secondary samples. Such mean centering often compensates for much of the spectral differences (eg, sloping background, wavelength shifts, and intensity differences) between primary and secondary samples, and in some cases, this centering is all that is actually needed.[23-25]

The linear system in Equations 6 and 7 has been previously proposed for model updating in chemometrics applications.[14,26,27] In other applications, the weight λ appears on the primary sample block[28]:

$$\begin{bmatrix} \lambda\mathbf{X}_P \\ \mathbf{X}_S \end{bmatrix} \mathbf{b} = \begin{bmatrix} \lambda\mathbf{y}_P \\ \mathbf{y}_S \end{bmatrix}, \tag{8}$$

Here, the λ value—between 0 and 1—plays the role of a "forgetting" factor whereby older primary samples in $\mathbf{X}_P$ are given less weight than more recent samples in $\mathbf{X}_S$.

Secondary predicting secondary and PPS will define the lower (better) and upper (worse) bounds of updating performance—the extremes where one calibrates on samples from only one domain (exclusively primary or secondary). Local mean centering will be the benchmark method that utilizes labeled samples from both the primary and secondary domains. The intent of this paper is to find updating methods whose performance is intermediate between SPS and LMC. Otherwise, LMC will simply suffice. The LMC method can be considered as a form of sample augmentation but to distinguish it from the new sample augmentation methods presented next, it is referred to as LMC.

3 | UPDATING VIA NULL SAMPLE AUGMENTATION

In this section, we will either add, via vertical concatenation, labeled secondary secondary samples $\{\mathbf{X}_S, \mathbf{y}_S\}$, unlabeled secondary spectra $\mathbf{X}_{SU}$, or a combination of both to the pool of primary samples $\{\mathbf{X}_P, \mathbf{y}_P\}$. The resulting model vector $\mathbf{b}$ depends upon two tuning parameters (k and τ). The methods are listed in Table 1.

Instead of adding labeled secondary samples $\{\mathbf{X}_S, \mathbf{y}_S\}$ to the pool of primary calibration samples $\{\mathbf{X}_P, \mathbf{y}_P\}$, we would ideally like to add only unlabeled secondary spectra $\mathbf{X}_{SU}$ for purposes of updating. Our framework for incorporating spectral information from both $\mathbf{X}_P$ and $\mathbf{X}_{SU}$ is expressed in the following linear system:

$$\min_{\mathbf{b}} \left(||\mathbf{X}_P\mathbf{b} - \mathbf{y}_P||_2^2 + \tau^2 ||\mathbf{R}\mathbf{b}||_2^2 \right) \iff \begin{bmatrix} \mathbf{X}_P \\ \tau\mathbf{R} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y}_P \\ \mathbf{0} \end{bmatrix}. \tag{9}$$

The matrix $\mathbf{R}$ has dimension $n_R \times d$. The number of rows n_R will depend upon the approach used to construct $\mathbf{R}$, and these approaches will be described in the next few subsections. The above linear system essentially approximates the following equality-constrained least squares problem: solve $\mathbf{X}_P\mathbf{b} = \mathbf{y}_P$ subject to the equality constraints[29,30] $\mathbf{R}\mathbf{b} = \mathbf{0}$. Mathematically, $\mathbf{R}\mathbf{b} = \mathbf{0}$ implies that the model vector $\mathbf{b} = [b_1, \ldots, b_d]^T$ of regression coefficients is orthogonal to $\mathbf{R}$. This sample augmentation constitutes a simple yet effective form of orthogonal signal correction: the regression model is attempting to be insensitive to undesirable spectral information contained in $\mathbf{R}$ by pointing the model vector $\mathbf{b}$ away from it. This observation forms the basis of many calibration maintenance and transfer methods (see other studies[14,15,27,31] and the references contained therein) and augmented classical least squares (ACLS)[17,32-35] procedures that decompose spectra into pure-component concentrations and pure-component spectra. Our approach differs from previous methods in that we include secondary samples with unknown reference measurements.

In some cases, the construction of such an $\mathbf{R}$ matrix is quite natural and spectroscopically intuitive. If one has a collection of nonanalyte spectra, then one can simply set $\mathbf{R}$ to be equal to this collection.[14,15,31,36,37] The interpretation here is

the following: we want our model vector to point away from this undesirable subspace spanned by spectra containing no analyte signals of interest. Note that in Equation 9, as $\tau \to \infty$, the model vector $\mathbf{b}$ approaches the net analyte solution $\mathbf{b}_{\text{NAS}}$

$$\mathbf{b}_{\text{NAS}} = \mathbf{X}_{\text{NAS}}^{+}\mathbf{y} \qquad \text{where} \qquad \mathbf{X}_{\text{NAS}} = \mathbf{X}_{\text{P}}\left(\mathbf{I} - \mathbf{V}_{\text{R}}\mathbf{V}_{\text{R}}^{\text{T}}\right), \tag{10}$$

where $\mathbf{V}_{\text{R}}$ are the loading vectors obtained from the singular value decomposition (SVD) of $\mathbf{R}$.[21,29,37] The pitfall of the net analyte solution approach is the following: too much analyte information is often removed in the null projection from $\mathbf{X}_{\text{P}}$ to $\mathbf{X}_{\text{NAS}}$.[27,37]

With respect to nonanalyte spectra, there are two special cases: difference spectra and constant-analyte spectra. If one has a fixed standardization set, then the samples can be measured across two instruments. Hence, two sets of spectra, $\mathbf{X}_1$ and $\mathbf{X}_2$ from the first and second instruments, respectively, are obtained. The difference spectra $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2$ embodies instrument-to-instrument variation that we want our model vector $\mathbf{b}$ to be insensitive to. This approach has been utilized in Kalivas et al.[14] In the case of a collection of constant-analyte spectra, one could similarly create difference spectra $\mathbf{R}$ by subtracting out the mean of the constant-analyte spectra from each spectrum in the collection. The subspace spanned by this $\mathbf{R}$ embodies sample-to-sample differences that we want to immunize our model against.

In the absence of difference or non-analyte spectra, the construction of an undesirable subspace spanned by $\mathbf{R}$ that is (*a*) spectroscopically meaningful and (*b*) uses only unlabeled secondary spectra is not self-evident. In the next few subsections, three approaches to creating a matrix $\mathbf{R}$ will be proposed. Since all of the methods require that the solution vector $\mathbf{b}$ lies in the null space of $\mathbf{R}$, ie, $\mathbf{Rb} = \mathbf{0}$, we will use the acronym NAR (short for null augmented regression) to collectively refer to the collection of calibration updating methods that differ only in their spectral content for $\mathbf{R}$. We want to ascertain how close in performance the NAR collection of methods (leveraging unlabeled secondary samples) is to benchmark performance of LMC (leveraging labeled secondary samples).

3.1 | NAR primary and NAR secondary

What follows are two formulations based upon null projections: null augmented regression secondary (NARS) and null augmented regression primary (NARP).

$$\mathbf{R} = \begin{cases} \mathbf{X}_{\text{P}}(\mathbf{I} - \mathbf{V}_{\text{SU}}\mathbf{V}_{\text{SU}}^{\text{T}}), & \text{NARP} \\ \mathbf{X}_{\text{SU}}(\mathbf{I} - \mathbf{V}_{\text{P}}\mathbf{V}_{\text{P}}^{\text{T}}), & \text{NARS} \end{cases}, \tag{11}$$

where $\mathbf{X}_{\text{P}}$ and $\mathbf{X}_{\text{SU}}$ are decomposed via their respective SVDs:

$$\mathbf{X}_{\text{P}} = \mathbf{U}_{\text{P}}\Sigma_{\text{P}}\mathbf{V}_{\text{P}}^{\text{T}} \quad \text{and} \quad \mathbf{X}_{\text{SU}} = \mathbf{U}_{\text{SU}}\Sigma_{\text{SU}}\mathbf{V}_{\text{SU}}^{\text{T}}. \tag{12}$$

For NARP, the matrix $\mathbf{R}$ has dimension $n_{\text{P}} \times d$ with the second block of Equation 9 being $\mathbf{Rb} = \mathbf{X}_{\text{P}}(\mathbf{I} - \mathbf{V}_{\text{SU}}\mathbf{V}_{\text{SU}}^{\text{T}})\mathbf{b} = \mathbf{0}$. The matrix $\mathbf{R}$ embodies that part of the primary spectra $\mathbf{X}_{\text{P}}$ that is insensitive to spectral variations in the unlabeled secondary spectra $\mathbf{X}_{\text{SU}}$. Moreover, the orthogonality mechanism of $\mathbf{Rb} = \mathbf{0}$ pushes $\mathbf{b}$ away from the undesirable subspace spanned by $\mathbf{R}$ and toward $\mathbf{X}_{\text{SU}}$, which is precisely the space we want to update to. For NARS, the matrix $\mathbf{R}$ has dimension $n_{\text{S}} \times d$ with $\mathbf{R} = \mathbf{X}_{\text{SU}}(\mathbf{I} - \mathbf{V}_{\text{P}}\mathbf{V}_{\text{P}}^{\text{T}})$ embodying spectral variations in $\mathbf{X}_{\text{SU}}$ that are not in $\mathbf{X}_{\text{P}}$. Here, the matrix $\mathbf{R}$ can be interpreted as the unique contribution of the secondary spectra relative the primary spectra. However, the model vector $\mathbf{b}$ is also orthogonal to this new contribution. Hence, when one predicts a new secondary spectrum $\mathbf{x}_{\text{S}}$ with this $\mathbf{b}$, only that component in $\mathbf{x}_{\text{S}}$ that resides in the space spanned by $\mathbf{X}_{\text{P}}$ contributes to the overall prediction.

Caution is advised whenever null orthogonal projections are employed. If there is significant interaction and overlap between the shared spectral interferents of $\mathbf{X}_{\text{P}}$ and $\mathbf{X}_{\text{S}}$ and the analyte, then there will be very little analyte information left after the projection. In this case, the use of orthogonal projections can be more deleterious performance-wise than the use of an oblique projections or no projections at all.[27] For both NARS and NARP, the crucial question is as follows: how much of the analyte information is lost in the null projection? If too much, then NARP and NARS will likely fail. A wavelength selection variation of NARS was used as a potential local modeling method, but the problem of too much lost analyte information persisted.[38]

For NARP, Equation 11 assumes that the loading vectors are derived from the unlabeled secondary samples. However, this membership requirement can easily be relaxed. For example, suppose one defines a general set of secondary spectra $\mathbf{X}_{\text{SG}}$ comprised of some combination of secondary samples from the calibration spectra $\mathbf{X}_{\text{S}}$, the unlabeled secondary spectra $\mathbf{X}_{\text{SU}}$ or (if available) the validation secondary spectra $\mathbf{X}_{\text{SV}}$, then one can define the residual matrix as

$$\mathbf{R} = \mathbf{X}_{\text{P}}(\mathbf{I} - \mathbf{V}_{\text{SG}}\mathbf{V}_{\text{SG}}^{\text{T}}) \quad \text{where} \quad \mathbf{X}_{\text{SG}} = \mathbf{U}_{\text{SG}}\Sigma_{\text{SG}}\mathbf{V}_{\text{SG}}^{\text{T}}, \tag{13}$$

via SVD. If $\mathbf{R}$ does involve loading vectors from the validation samples $\mathbf{X}_{SV}$, then NARP via Equation 9 can be viewed as a form of transductive regression.[39] In this transductive case, we are using the additional wavelengths/wave numbers to glean additional sample density information that may help with prediction.[40] For convenience, we will not consider all possible subsets of $\{\mathbf{X}_S, \mathbf{X}_{SU}, \mathbf{X}_{SV}\}$ but just $\mathbf{X}_{SG} = \mathbf{X}_{SU}$, which simplifies to NARP and NARS.

3.2 | NAR eigenvalue

Generalized eigenvalue problems (GEPs) form another way to construct an undesirable subspace $\mathbf{R}$. Generalized eigenvalue problems naturally arise in many diverse fields such as signal processing, pattern recognition, and machine learning.[41-43] These problems involve data from two different domains: *source* and *target* (or primary and secondary, respectively, in the language of chemometrics). Using the notation introduced in this paper, GEPs typically involve two matrices: total scatter $\mathbf{T}$ and domain scatter $\mathbf{D}$:

$$\mathbf{T} = \mathbf{X}_c^T \mathbf{X}_c \quad \text{and} \quad \mathbf{D} = (\boldsymbol{\mu}_P - \boldsymbol{\mu}_{SU})(\boldsymbol{\mu}_P - \boldsymbol{\mu}_{SU})^T, \tag{14}$$

such that $\mathbf{X} = [\mathbf{X}_P; \mathbf{X}_{SU}]$ is the vertical stacking of spectral domains, and $\mathbf{X}_c = \mathbf{X} - \mathbf{1}_n \boldsymbol{\mu}_X^T$ is the mean-centered version of $\mathbf{X}$ with $n = n_P + n_{SU}$. Here, the spectral means of $\mathbf{X}$ and $\mathbf{X}_{SU}$ are specified as follows:

$$\boldsymbol{\mu}_X = \frac{1}{n}\mathbf{X}^T\mathbf{1}_n \quad \text{and} \quad \boldsymbol{\mu}_{SU} = \frac{1}{n_{SU}}\mathbf{X}_{SU}^T\mathbf{1}_{n_{SU}}. \tag{15}$$

Note that in linear discriminant analysis, one objective is to maximize domain (or between-class) scatter, ie, find a model vector $\mathbf{b}$ that makes the difference between domain centroids as large as possible. Here, we seek the opposite: find vectors $\mathbf{b}$ that minimize this difference.

For calibration updating purposes, the matrices $\mathbf{T}$ and $\mathbf{D}$ typically occur in the following minimization problem:

$$\min_{\mathbf{b}} \mathbf{b}^T\mathbf{D}\mathbf{b} \quad \text{subject to} \quad \mathbf{b}^T\mathbf{T}\mathbf{b} = 1. \tag{16}$$

By minimizing $\mathbf{b}^T\mathbf{D}\mathbf{b}$, we explicitly seek model vectors $\mathbf{b}$ that minimize domain scatter that pull together centroids from primary and (unlabeled) secondary spectra. The inner product $\mathbf{b}^T\mathbf{T}\mathbf{b}$ reflects total scatter and is analogous to principal component analysis: find vectors $\mathbf{b}$ of unit length such that variance is maximized. Using the standard approach of Lagrange multipliers, we obtain the following GEP[44,45]:

$$\mathbf{D}\mathbf{b} = \delta\mathbf{T}\mathbf{b} \;\Rightarrow\; \mathbf{R}\mathbf{b} = \mathbf{0} \text{ where } \mathbf{R} = \mathbf{D} - \delta\mathbf{T}. \tag{17}$$

The solutions to $\mathbf{R}\mathbf{b} = \mathbf{0}$ correspond to generalized eigenvectors that minimize domain scatter and maximize total scatter. The scalar δ corresponds to a generalized eigenvalue, and it takes on two unique nonnegative values: $\delta = \{0, \delta_*\}$. Here, $\delta_* > 0$ is a nonzero eigenvalue that occurs only once while $\delta = 0$ is a repeated eigenvalue. Based upon the performance observations in Andries[46] and Ghifary,[47] scatter matrices $\mathbf{R}$ where $\delta = 0$ are generally noninferior to the cases where δ is nonzero in value, and we will use $\delta = 0$ as the default value. When $\delta = 0$, $\mathbf{R}$ simplifies to just the domain scatter matrix: $\mathbf{R} = \mathbf{D} = (\mu_P - \mu_{SU})(\mu_P - \mu_{SU})^T$. In fact, since $\mathbf{D}$ is a rank-one matrix, $\mathbf{R}$ can be further simplified to just one row: $\mathbf{R} = (\mu_P - \mu_{SU})^T$. We will use the acronym NARE (short for NAR eigenvalue) to this approach of minimizing domain scatter.

4 | UPDATING VIA FEATURE AUGMENTATION

The previously introduced approaches have been premised on the following idea: that one regression vector alone is enough to perform calibration updating. However, multicomponent regression methods have been proposed.[48-50] This section introduces methods based upon this concept. As noted previously, the methods introduced in this section are summarized in Table 1 with PLS being the solution approach.

4.1 | Example of feature augmentation

In an example from Google, internet viewing data from two sources are combined: a small set of expensive high-quality observations (analogous to the secondary samples $\{\mathbf{X}_S, \mathbf{y}_S\}$) and a much larger set with less costly observations (analogous to the primary samples $\{\mathbf{X}_P, \mathbf{y}_P\}$).[49] The goal is to make predictions of internet viewing habits for samples drawn from

the smaller set. Their proposed least squares minimization problem is the following (using our analogous chemometrics notation):

$$||\mathbf{X}_S\mathbf{b}_S - \mathbf{y}_S||_2^2 + ||\mathbf{X}_P(\mathbf{b}_S + \mathbf{b}_P) - \mathbf{y}_P||_2^2 + \lambda^2\,||\mathbf{X}_S\mathbf{b}_P||_2^2 \iff \begin{bmatrix} \mathbf{X}_P & \mathbf{X}_P \\ \mathbf{0} & \mathbf{X}_S \\ \lambda\mathbf{X}_S & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b}_P \\ \mathbf{b}_S \end{bmatrix} = \begin{bmatrix} \mathbf{y}_P \\ \mathbf{y}_S \\ \mathbf{0} \end{bmatrix}. \tag{18}$$

The boldface zeros indicate matrices and vectors of appropriate dimension such that the matrix/vector concatenations make sense. Here, the vector $\mathbf{b}_P$ can be thought of as a perturbation of the model vector $\mathbf{b}_S$. The prediction on a new primary sample $\mathbf{x}_P$ is then based upon the sum $\mathbf{b} = \mathbf{b}_P + \mathbf{b}_S$ where $\hat{y}_P = \left(\mathbf{x}_P - \boldsymbol{\mu}_P^X\right)^T\mathbf{b} + \mu_P^y$.

4.2 | Feature augmentation via LMC

We propose a modification of Equation 18 but based upon the LMC approach in Equation 7:

$$||\mathbf{X}_P\mathbf{b}_P - \mathbf{y}_P||_2^2 + \lambda^2\,||\mathbf{X}_S(\mathbf{b}_P + \mathbf{b}_S) - \mathbf{y}_S||_2^2 \iff \begin{bmatrix} \mathbf{X}_P & \mathbf{0} \\ \lambda\mathbf{X}_S & \lambda\mathbf{X}_S \end{bmatrix} \begin{bmatrix} \mathbf{b}_P \\ \mathbf{b}_S \end{bmatrix} = \begin{bmatrix} \mathbf{y}_P \\ \lambda\mathbf{y}_S \end{bmatrix}. \tag{19}$$

Equation 19 will be denoted as FA-1. The prediction on a new secondary spectrum $\mathbf{x}_S$ is also based upon the sum of model vectors $\mathbf{b} = \mathbf{b}_P + \mathbf{b}_S$ such that $\hat{y}_S = \left(\mathbf{x}_S - \boldsymbol{\mu}_S^X\right)^T\mathbf{b} + \mu_S^y$. Unlike Equation 18, $\mathbf{b}_S$ is a perturbation of the model vector $\mathbf{b}_P$. In Equation 19, we may want to add an additional constraint: the secondary model vector $\mathbf{b}_S$ is insensitive to spectral variations in the primary spectra $\mathbf{X}_P$. That is, we want $\mathbf{b}_S$ to point away from this undesirable subspace, and as a consequence, we arrive at the orthogonality constraint $\mathbf{X}_P\mathbf{b}_S = \mathbf{0}$. To accommodate this additional restriction, we can augment with another block in a similar fashion to Equation 18:

$$\begin{bmatrix} \mathbf{X}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_P \\ \lambda\mathbf{X}_S & \lambda\mathbf{X}_S \end{bmatrix} \begin{bmatrix} \mathbf{b}_P \\ \mathbf{b}_S \end{bmatrix} = \begin{bmatrix} \mathbf{y}_P \\ \mathbf{0} \\ \lambda\mathbf{y}_S \end{bmatrix}. \tag{20}$$

Equations 19 and 20 will be denoted as FA-1 and FA-2, respectively, where FA is an acronym for feature augmentation.

5 | EXPERIMENTAL

5.1 | Data sets

We examine two NIR data sets. One measures Juniper berry concentration in fecal spectra, and another measures moisture, protein, and oil concentrations from soy seeds.

5.1.1 | Goat feces (GOAT)

One hundred nine (109) NIR spectra were measured from 400 to 2498 nm at 2-nm intervals (1050 wavelengths).[51,52] The analyte of interest is Juniper berry concentration (%w/w) from goat fecal spectra. The samples are categorized into two feeding trials with each trial corresponding to a different year: 1999 (61 samples) and 2002 (48 samples). Primary and secondary samples correspond to samples drawn from years 1999 and 2002, respectively.

5.1.2 | Soy seed (SOYS)

NIR spectra from 60 soy seed samples were collected in reflectance mode from 1102 to 2498 nm at 4-nm intervals (350 wavelengths) on two NIR instruments.[53,54] Spectra were the average of five replicates. Reference values for moisture, protein, and oil were determined by reference methods from wet chemistry. This data set has a standardization set that can be defined, but such a set is not used in this study. Recall that a standardization set, as defined in this paper, is a set of samples measured across both the primary and secondary conditions. This distinction matters when we form multiple calibration and validation sets via random shuffling in section 5.4, so as not to incorporate the standard set scenario.

5.2 | Sample sets

Depending upon the calibration updating method, different samples from the primary and secondary data will be used to build the calibration model. The number of samples used in the sample sets for GOAT and SOYS are given in Table 2.

TABLE 2 For each data set, the number of samples used in each sample block[a]

	CALP $\{\mathbf{X_P}, \mathbf{y_P}\}$	CALS $\{\mathbf{X_S}, \mathbf{y_S}\}$	UNLS $\mathbf{X_{SU}}$	VALS $\{\mathbf{X_{SV}}, \mathbf{y_{SV}}\}$	Total
SOYS	30	3 (6)	15	9	60
GOAT	61	5 (10)	24	14	109

[a]Values in parentheses () correspond to the case where the number of samples in the CALS sample set is doubled.

No attempt is made to optimize the selection of calibration samples (eg, Kennard Stone algorithm). We want to mimic what one often encounters in practice (especially if one is not involved in the design of experiment): a *given* calibration set consisting of primary and secondary samples that are distinct. We will now define these sample sets.

Primary samples: The primary calibration sample set $\{\mathbf{X}_P, \mathbf{y}_P\}$ will be denoted by the acronym CALP. For GOAT, all 61 samples from year 1999 are considered primary calibration samples. For SOYS, the first 50% of the samples from the first instrument are deemed calibration samples. The actual number of respective CALP samples are listed in Table 2.
Secondary samples in GOAT: All samples from year 2002 are secondary samples. These samples are subdivided into three sets: a calibration secondary set (CALS or $\{\mathbf{X}_S, \mathbf{y}_S\}$), an unlabeled secondary set (UNLS or $\mathbf{X}_{SU}$) whereby the reference measurements for the spectra in $\mathbf{X}_{SU}$ are intentionally masked, and a validation secondary set (VALS or $\{\mathbf{X}_{SV}, \mathbf{y}_{SV}\}$). The CALS and UNLS sample sets are used for calibration while the VALS set is used purely for performance assessment (and never for calibration/training). The sample percentages in CALS, UNLS, and VALS are 10%, 50%, and 30%, respectively. The percentages 10%, 50%, and 30% account for 90% of the secondary samples. The remaining 10% will be set aside for purposes of performance assessment when the percentage of CALS samples is enlarged from 10% to 20%. Specifically, in the default case, we will assess performance when the CALS set is small, ie, 10% of the secondary samples. We will then double the size of $\{\mathbf{X}_S, \mathbf{y}_S\}$ from 10% to 20%.
Secondary samples in SOYS: Only the second 50% of samples from the second instrument will be set aside for the CALS, UNLS, and VALS sample sets (preserving the same percentages—10%, 50%, and 30% that was used in GOAT). The first 50% of the samples from the second instrument is ignored (except for SPS, the reason for which is described in section 5.3). We want to mimic what one observes in GOAT: a separate and nonoverlapping set of primary (first instrument) and secondary (second instrument) samples. For example, there are 60 samples measured on both instruments. The first 30 samples (50%) on the first instrument (indices 1:30) are the CALP samples. The next 30 samples on the second instrument (indices 31:60) will be used as the secondary samples. Index sets 31:36, 37:51, and 52:60 correspond to the CALS, UNLS, and VALS sample sets. (Indices 34:36 from the second instrument will be used to enlarge the CALS sample set from 10% to 20%.)

5.3 | Calibration methods and their corresponding sample sets

For PPS, only the CALP samples $\{\mathbf{X}_P, \mathbf{y}_P\}$ are used to build the primary calibration model. For SSPS, only the CALS set comprises the set of calibration secondary samples $\{\mathbf{X}_S, \mathbf{y}_S\}$. For GOAT, to get a large enough quorum of secondary samples to achieve the optimal level of SPS performance, the CALS and UNLS sample sets are combined to form the SPS calibration model. Forming this secondary calibration model for SPS is the only time the reference measurements for the UNLS sample set are unmasked. For SOYS, we combine the CALS and UNLS sample sets with the first 50% of the samples from the second instrument (that are otherwise ignored) to get a large enough quorum of secondary samples for SPS.

The NAR collection of methods (NARP, NARS, and NARE) use both the CALP samples $\{\mathbf{X}_P, \mathbf{y}_P\}$ and UNLS spectra $\mathbf{X}_{SU}$ for calibration. Note that the sample sets in these unlabeled methods are assembled in the inductive mode (the UNLS set is simply $\mathbf{X}_{SU}$) rather than in the transductive mode (the UNLS set being enlarged to $\mathbf{X}_{SU} \cup \mathbf{X}_{SV}$). The benchmark method LMC and the new FA methods FA-1 and FA-2 use both the CALP samples $\{\mathbf{X}_P, \mathbf{y}_P\}$ and the CALS samples $\{\mathbf{X}_S, \mathbf{y}_S\}$ for calibration.

5.4 | Data splits

Performance metrics (to be discussed in section 5.6) will be measured across 300 random splits of the data. Performance results over one fixed split of the data are often anecdotal. By using many shuffled splits of the data, we intend to examine

the sensitivity and robustness of the updating methods to sample perturbations. The sample set proportions for both data sets (previously described and overviewed in Table 2) are maintained on each of the 300 random splits.

5.5 | Tuning parameters

Each updating method has either one, two, or three tuning parameters associated with it. The parameters associated with each updating method is as follows. Primary predicting secondary, SPS, and SSPS have one tuning parameter k, the number of PLS latent vectors ("PLS factors" or "PLS LVs"). Local mean centering, FA-1, and FA-2 have two tuning parameters (k, λ) while the NAR collection of methods (NARP, NARS, and NARE) have two tuning parameters (k, τ).

The number of PLS latent vectors k will range across 30 values: $k = 1, 2, \ldots, 30$. The penalty parameter λ will span 50 values in an exponentially decaying fashion: $\lambda_1 > \lambda > \cdots > \lambda_{50}$. The value λ_1 corresponds to the largest singular value of $\mathbf{X}_P$ (or $\mathbf{X}_S$ in the case of SPS or SSPS) and λ_{50} is half the value of the smallest nonzero singular value of $\mathbf{X}_P$ (or $\mathbf{X}_S$). This exponentially decaying sampling strategy is based upon the recommendation in Hansen.[29] We do not want to fix a set of λ values across both data sets since a given λ value may overregularize (smooth too much) or underregularize (smooth too little) model vectors $\mathbf{b}$ from different data sets. The τ values will also span the same set of λ values. It should be noted that NARS and NARP rely on the number of loading vectors used to form $\mathbf{V}_{SU}$ and $\mathbf{V}_P$ for respective projections. In these situations, to simplify the selection of the number of loading vectors used, all $r_{SU} - 1$ loading vectors are used where r_{SU} is the rank of $\mathbf{X}_{SU}$. Similarly for $\mathbf{X}_P$, all $r_P - 1$ loading vectors are used where r_P is the rank of $\mathbf{X}_P$. A model vector will be generated for each set of parameters values: 30 model vectors for PPS, SPS, and SSPS and $1500 = (30)(50)$ model vectors for the NAR and FA collections of methods.

The justification for using $r_P - 1$ and $r_{SU} - 1$ loading vectors for NARP and NARS, respectively, are as follows. In any situation where the SVD is used to form a projection matrix, eigenvector selection occurs, ie, how many of the full rank r_P or r_{SU} eigenvectors should be used? For example, popular orthogonal projection methods for net analyte signal (NAS) preprocessing are noted as requiring eigenvector selection. Methods have been developed to address eigenvector selection; see Vitale et al[55] and Malinowski[56] and references therein. Such eigenvector selection methods could be used for NARP and NARS or a simple rule that eigenvectors are included up to 99% of the information.[21] Alternatively, the number of eigenvectors can be an additional model selection tuning parameter as implemented in recent work with a NARS-type model updating process.[38] However, past model updating work indicated that a small sample subset of full spectral differences between primary and secondary performs better than LMC (as presented in this paper that utilizes a small subset of secondary spectra and references values).[14] Thus, a small study was undertaken for this paper to assess the degree of changes in the NARP and NARS $\mathbf{R}$ matrices upon each addition of an eigenvector to form the respective projection matrices. We created SOYS heat maps (not shown) for successive NARP $\mathbf{R}$ matrices representing respective correlations between the rows (samples) and columns (wavelengths) of $\mathbf{R}$. (Similar heat map trends were observed for GOAT for NARS and NARP.) These heat maps show erratic correlation behavior at early eigenvectors. When the correlations—between the $\mathbf{R}$ matrix constructed from the first i eigenvectors and the $\mathbf{R}$ matrix constructed from the first $(i + 1)$ eigenvectors—converge, the rows (the projected spectra making up $\mathbf{R}$) are not significantly changing, and it is assumed that the full orthogonal spectral differences between the primary and secondary conditions has been captured. (It should be noted that this is the full orthogonal spectral differences and not the spectral differences based on the same set of samples being measured in both conditions.) Since we observed that the correlations between successive $\mathbf{R}$ matrices were rapidly changing as i increased from low to intermediate values, it would be difficult to fine tune eigenvector selection to just the right subset. It was found that the convergence typically occurs in the last 10% of eigenvectors (beyond the 99% information level or variance explained). Thus, for this work, all eigenvectors up to $r_P - 1$ and $r_{SU} - 1$ are used to form the NARP and NARS projection matrices, respectively, in this paper.

5.6 | Performance metrics

Two performance metrics will be of interest: root mean square error of validation (RMSEV) and R^2 of validation. For a given model, the RMSEV is computed as

$$\text{RMSE}_{\text{VAL}} = \sqrt{\frac{1}{n_{\text{SV}}} \sum_{i=1}^{n_{\text{SV}}} \left(\hat{y}_{(\text{SV,i})} - y_{(\text{SV,i})} \right)^2}, \tag{21}$$

where n_{SV} indicates the number of samples in the VALS sample set. The R^2 value is the coefficient of determination associated with the regression line that fits the scatterplot associated with reference values (x-axis) and prediction estimates (y-axis). The RMSEV and R^2 values will be computed for each tuning parameter set. The following subscripted j indicates the performance metric computed on the jth model: RMSEV_j and R^2_j. For example, the LMC performance on a given data set would consist of 1500 models, ie, the jth model corresponds to a tuning parameter pair $(k_{i_k}, \lambda_{i_\lambda})$, $i_k = 1, \ldots, 30$, $i_\lambda = 1, \ldots, 50$.

We are interested in assessing performance trends across the calibration updating methods. For each updating method and for each data split, we compute the lower quartile (minimum, 25th percentile and median) RMSE_{VAL} values across all models. Using LMC again as an example, we have 1500 RMSE_{VAL} values for each of the 300 data splits. For performance assessment purposes, all of this gets distilled to a 300×3 matrix $\mathbf{Q}$ of lower quartile values:

$$\mathbf{Q}_{\text{RMSE}} = \begin{bmatrix} Q_{(1,0)} & Q_{(1,1)} & Q_{(1,2)} \\ Q_{(2,0)} & Q_{(2,1)} & Q_{(2,2)} \\ \vdots & \vdots & \vdots \\ Q_{(300,0)} & Q_{(300,1)} & Q_{(300,2)} \end{bmatrix}. \tag{22}$$

The first, second, and third columns of $\mathbf{Q}_{\text{RMSE}}$ correspond to the minimum, 25th percentile (first quartile), and median (second quartile) RMSEV values, respectively, and each row corresponds to a data split. For example, the $(i, 1)$ entry of $\mathbf{Q}_{\text{RMSE}}$ denotes the minimum RMSEV value across all 1500 LMC models for the ith data split. We will then examine the boxplot based upon each column of $\mathbf{Q}_{\text{RMSE}}$. Each updating method will have their own matrix $\mathbf{Q}_{\text{RMSE}}$ of RMSEV values and three corresponding boxplots.

The lower quartile values of each calibration updating method will be compared against the lower quartile values of PPS (the "no updating" scenario). The reason for this is threefold. First, for a given calibration updating method, we want to see if the bulk of its median RMSEV values fall below the bulk of the minimum RMSEV values for PPS. If so, then we have confidence in the success of that particular updating method regardless of model selection. (The model selection procedure that is eventually employed just has to outperform the median model that separates the bottom half [the "best" models] from the top half [the "worst" models] with respect to RMSEV.) Second, we want to see how the performance trends for one data set trends hold up against another data set. Does one updating method perform well on one data set but not another? Third, for a given updating method, we want to observe how close the minimum, 25th percentile, and median RMSEV values are. One could reasonably expect that the parameters associated with the minimum RMSEV values are overfitted. However, if the minimum, 25th percentile, and median RMSEV values are still close, then this observation builds additional confidence for the success of a particular updating method.

For each model associated with an entry in the lower quartile matrix $\mathbf{Q}_{\text{RMSE}}$, we find its corresponding R^2 value and similarly form a 300×3 matrix (denoted $\mathbf{Q}_{\text{R2}}$) of R^2 values. As with the RMSEV values, we will form a boxplot based upon each column of $\mathbf{Q}_{\text{R2}}$.

The focus of the paper is to report on performance trends across multiple data splits of the data sets. It should be noted that no statistical testing was performed between the methods. Statistical testing requires distribution assumptions and confidence level decisions, and we have chosen not to enter this foray.

6 | RESULTS

6.1 | Description of performance figures

Figure 1 gives the overall performance assessment for the corn data set with moisture as the analyte. Legend-wise, we will now discuss what is displayed in Figure 1. The following discussion of the figure description will hold for the remaining performance figures.

The rows associated with different background colors corresponds to R^2 values (top row) and the corresponding RMSEV values (second and third rows). The third row is the same as the second row except that the third row displays a limited range in the y-axis—the solid white line in the second row (at a RMSEV height slightly above 3) corresponds to the maximal height in the third row. The third row is meant to give greater visual detail at the lower RMSEV range. For the second and third rows, each updating method has three boxplots: left, middle, and right. These three boxplots correspond to the minimum, first quartile, and median RMSEV values, respectively.

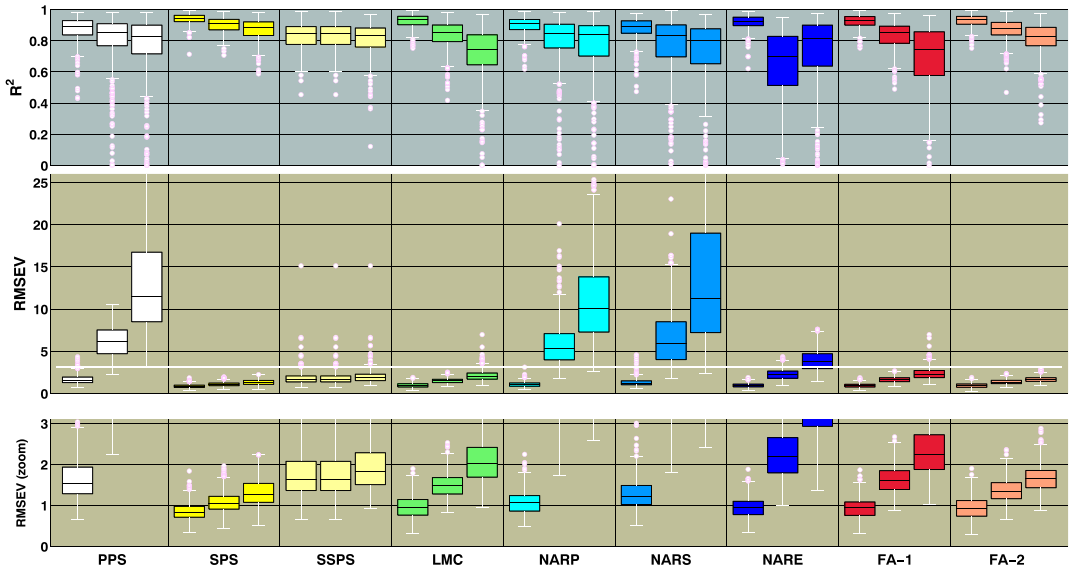

FIGURE 1 Performance boxplots for **SOYS moisture** across calibration updating methods. Here, 10% of the secondary samples were used for the CALS sample set. For each method in the second and third rows, the three boxplots correspond to the minimum, first quartile, and median RMSEV values. The boxplots in the top row are the corresponding R^2 values. CALS, calibration secondary set; FA, feature augmentation; LMC, local mean centering; NARE, null augmented regression eigenvalue; NARP, null augmented regression primary; NARS, null augmented regression secondary; PPS, primary predicting secondary; RMSEV, root mean square error of validation; SPS, secondary predicting secondary; SSPS, small secondary predicting secondary

6.2 | Performance trends

The various calibration updating methods examined here are divided into three classes. The first class consists of four benchmark methods: PPS, SPS, SSPS, and LMC. The NAR collection of methods (NARP, NARS, and NARE) comprise the second class. The third class of methods are the feature augmentation (FA) methods (FA-1 and FA-2). We like to remind the reader that the NAR collection does not use labeled secondary samples—the unlabeled secondary spectra $\mathbf{X}_{SU}$ are used instead.

6.2.1 | SOYS

Figure 1 reveals traits with moisture being the analyte of interest. The traits are shared by the other analytes of SOYS—see Figures 2 and 3 for oil and protein, respectively. These shared traits are now discussed.

From the PPS results, some form of model updating is needed. Again, the performance targets for the proposed methods are the SPS results. With respect to updating methods that use the labeled secondary samples, the benchmark method LMC and the FA methods perform comparably with FA-2 providing improved results compared to LMC with respect to low RMSEV and high R^2 behavior. The improvement of FA-2 over FA-1 shows the value of including the orthogonality constraint $\mathbf{X}_P\mathbf{b}_S = \mathbf{0}$ (see Equation 20). All of the updating methods with labeled secondary samples improve on SSPS, showing that the primary samples are needed. Aside from SSPS, increasing the number of secondary calibration for $\{\mathbf{X}_S, \mathbf{y}_S\}$ in CALS from 10% to 20% marginally improves performance for some methods, but no meaningful improvement is observed for most methods—see Figure 4A, 4B, and 4C. The lack in improvement does not appear to justify the doubling of the number of labeled samples (with respect to the expense and laboratory time associated with acquiring these additional reference measurements).

With respect to the updating methods that use only the unlabeled secondary spectra $\mathbf{X}_{SU}$, the NAR collection generally suffer from higher RMSEV values. However, what is particularly striking is how comparable the best quartile of NARE models (ie, the second dark-blue boxplot in the second and third rows) are relative to the best quartile of models for the LMC and FA methods. This is quite encouraging since only row—the spectral difference between $\mathbf{X}_P$ and $\mathbf{X}_{SU}$ centroids—was used in the updating (albeit each centroid is a mean of many samples). It is important to note that, across all data sets, only 50% of the secondary samples were set aside for $\mathbf{X}_{SU}$. However, these are small in terms of absolute numbers (see Table 2). In many applications, one could easily acquire hundreds or thousands of secondary samples without reference measurements. In such a case, it is possible that an $\mathbf{R}$ matrix constructed from many more unlabeled

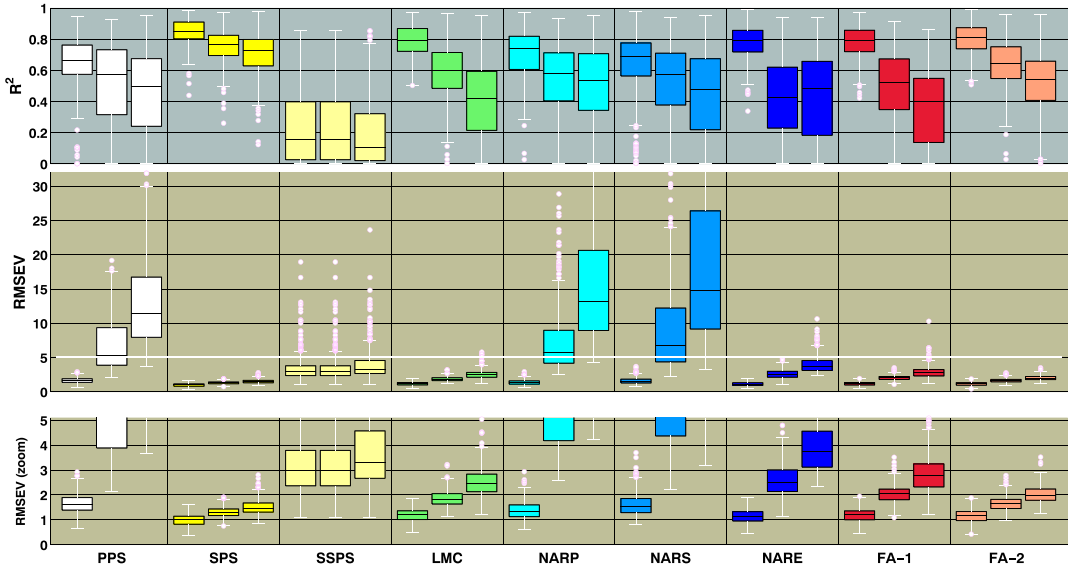

FIGURE 2 Performance boxplots for **SOYS oil** across calibration updating methods. Here, 10% of the secondary samples were used for the CALS sample set. See section 6.1 and Figure 1 for the description

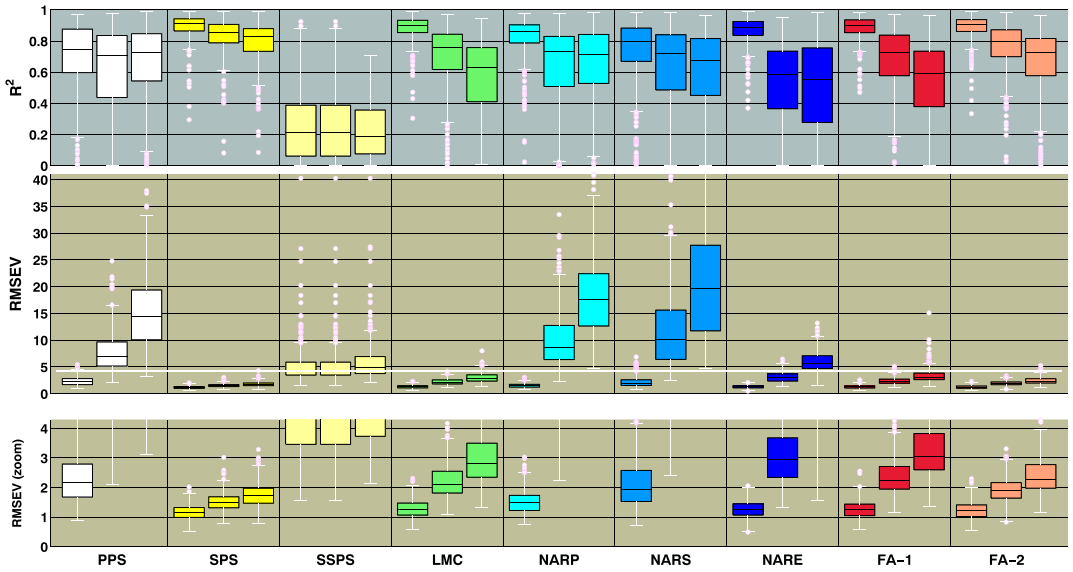

FIGURE 3 Performance boxplots for **SOYS protein** across calibration updating methods. Here, 10% of the secondary samples were used for the CALS sample set. See section 6.1 and Figure 1 for the description

secondary samples would better characterize the spectral variations associated with the secondary samples, thereby lowering RMSEV values.

One can sample augment the labeled secondary set $\{\mathbf{X}_S, \mathbf{y}_S\}$ to the linear system associated with the NAR methods:

$$\min_{\mathbf{b}} \left(||\mathbf{X}_P\mathbf{b} - \mathbf{y}_P||_2^2 + \lambda^2 ||\mathbf{X}_S\mathbf{b} - \mathbf{y}_S||_2^2 + \tau^2 ||\mathbf{R}\mathbf{b}||_2^2 \right) \quad \Longleftrightarrow \quad \begin{bmatrix} \mathbf{X}_P \\ \lambda\mathbf{X}_S \\ \tau\mathbf{R} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y}_P \\ \lambda\mathbf{y}_S \\ \mathbf{0} \end{bmatrix}. \tag{23}$$

The results (not shown) qualitatively mimic the performance of LMC. This demonstrates the power of including just a few reference samples. The information provided by the unlabeled secondary spectra—encoded by the matrix $\mathbf{R}$—is effectively ignored in favor of the labeled secondary samples $\{\mathbf{X}_S, \mathbf{y}_S\}$.

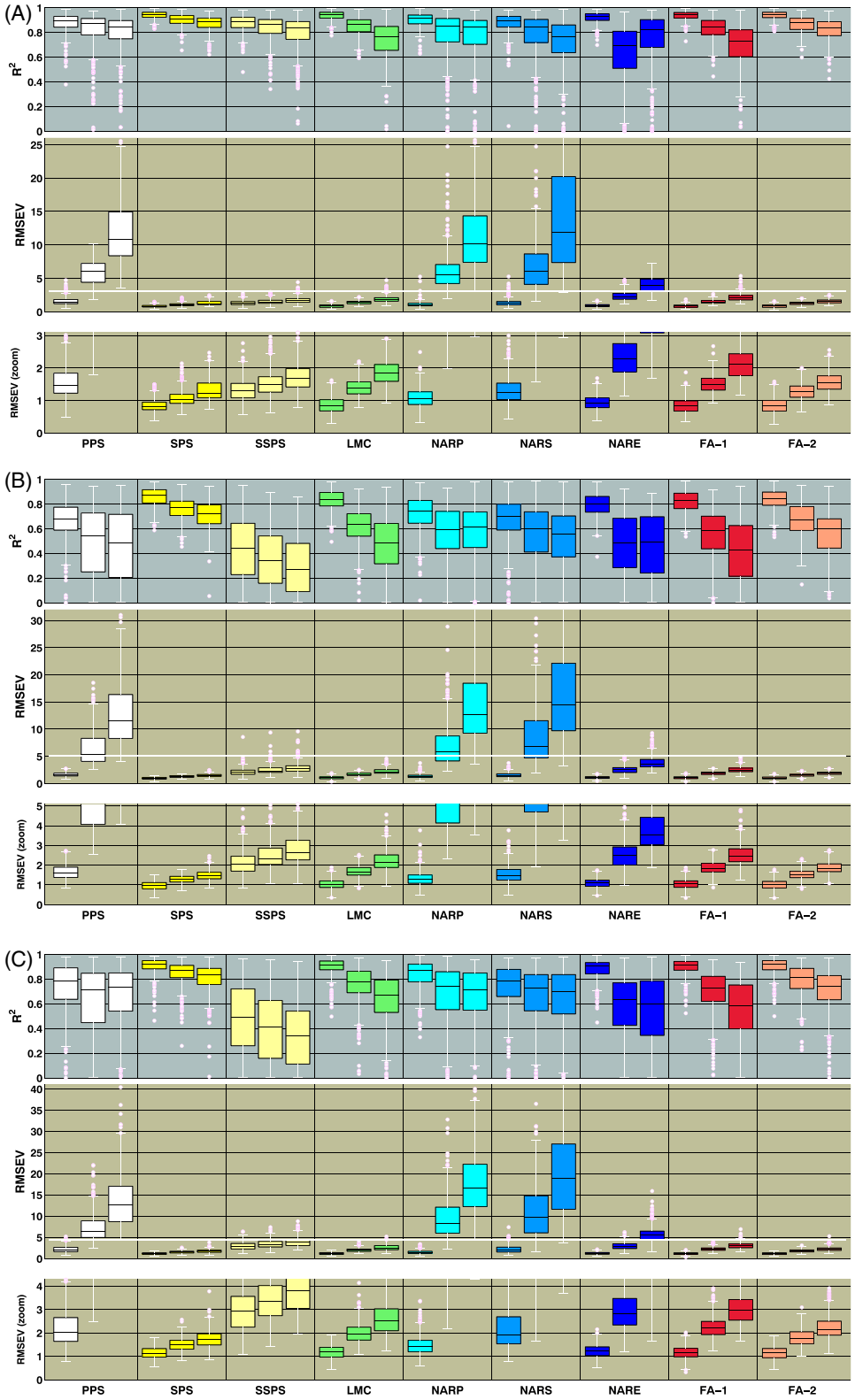

FIGURE 4 Performance boxplots for SOYS; A, moisture; B, oil; and C, protein across calibration updating methods corresponding to the 20% CALS sample set. See section 6.1 and Figure 1 for the description

Figure 5 displays the median RMSEV heatmaps for LMC, FA-2, and NARE, the exemplars for the benchmark, FA and NAR collection of methods. Note that Figure 5 displays the results for the 10% CALS sample set . (The heatmap for the 20% CALS sample set is not shown since the results are qualitatively similar.) Median RMSEV values for LMC and FA-2 span across $\log_{10}(\lambda)$ and k values while for NARE, the RMSEV values span across $\log_{10}(\tau)$ and k values. As viewed from the median perspective, LMC and FA-2 are robust across a large swath of tuning parameters. Null augmented regression eigenvalue, on other other hand, requires a large τ penalty—the weight on the **R** matrix in Equation 9—to be effective.

The median (Figure 5) has a large number of models with RMSEV around 1.2 and 2, with a good chance such a model could be selected. For Figure 6, the minimum, most of the models range between 0.29 and 1 RMSEV values indicating we also have a good chance of finding models in this lower range, perhaps lower than what the median heatmap suggests. If one develops model selection criteria such that near-optimal models are always chosen (perhaps an over-optimistic assumption), then Figure 6 is perhaps a more appropriate heatmap. As expected, the first quartile heatmap have lower RMSEV ranges, intermediate between the median and minimum heatmaps (not shown). The behavior for SOYS oil and protein are qualitatively similar.

For Figures 5 and 6, there is convergence at lower values of λ for LMC and NARE. For LMC, note the horizontal uniformity of the heatmap within the white black boxes on the far right. In these bounding-box convergence regions, RMSEV depends solely on the number of latent vectors k. (Since the color bar includes the higher RMSEV values associated with NARE, the heatmap resolution for LMC and FA-2 for the SOYS data set is not as fine-grained.) For FA-2, there is quasi-convergence. The convergences at small λ values is a normal trend, and essentially, the values of λ need to be small for modeling purposes. For NARE, and unlike LMC and FA-2, lower RMSEVs are obtained for larger values of λ. A larger weight is likely needed since **R** consists of only one row. From a regularization perspective, NARE tends to favor model vectors that are oversmoothed.

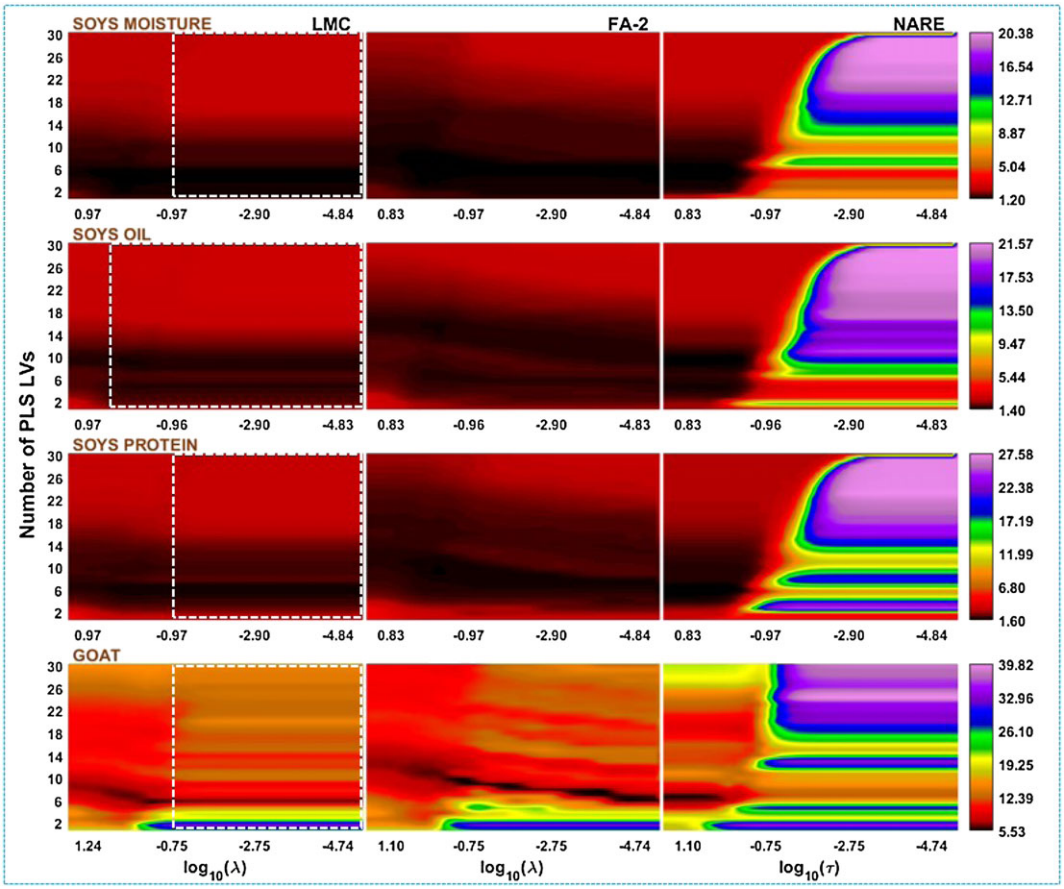

FIGURE 5 Heatmaps of median RMSEV values across tuning parameters: λ versus k for LMC (column 1) and FA-2 (column 2) and τ versus k for NARE (column 3). Each row corresponds to a data set (and for rows 2, 3, and 4, the corresponding analyte). Here, 10% of the secondary samples were used for the CALS sample set. The white dotted lines indicate where RMSEV is invariant with respect to the value of λ

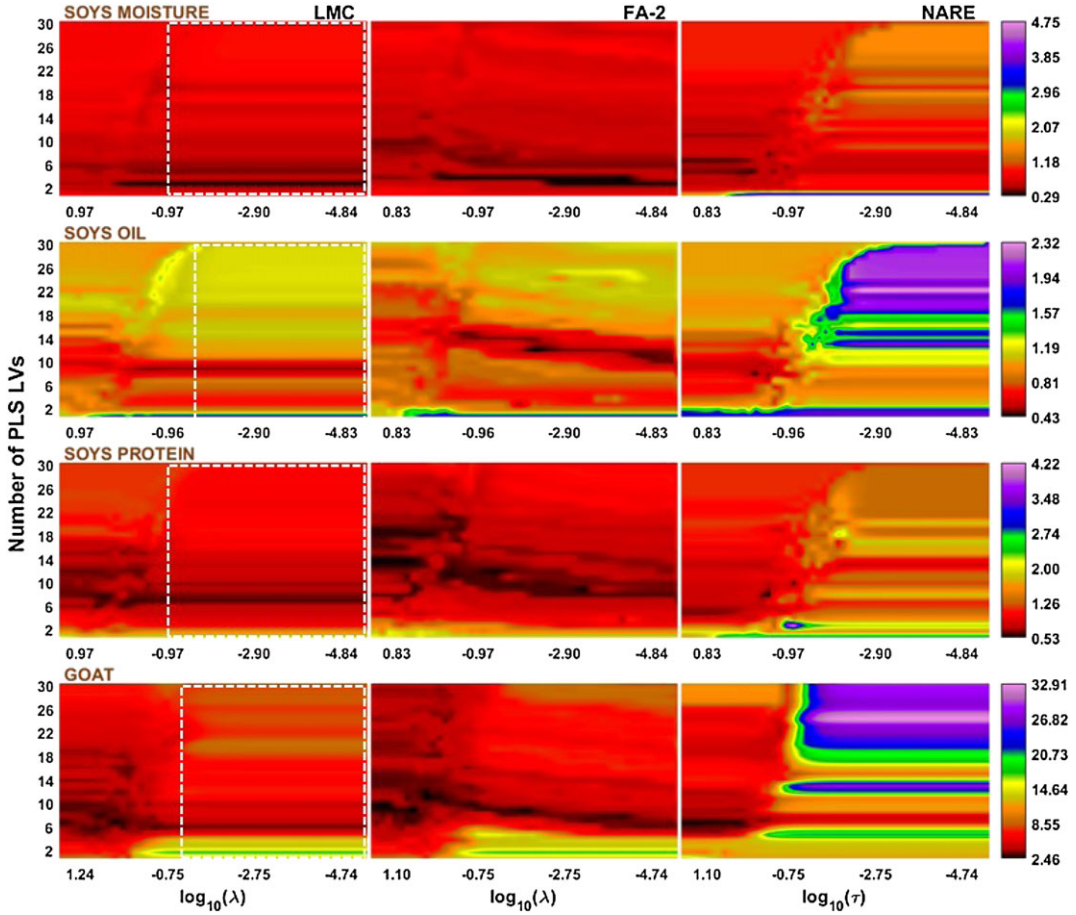

FIGURE 6 Heatmaps of minimum RMSEV values across tuning parameters: λ versus k for LMC and FA-2 and τ versus k for NARE

With respect to the performance of LMC and FA-2, two questions arise:

1. How often does LMC outperform and underperform FA-2 across all splits?
2. When the benchmark method LMC does outperform (underperform) FA-2, by how much, on average, does LMC outperform (underperform) with respect to RMSEV?

To answer these questions, the 300 minimum RMSEV values used to form the respective boxplots for LMC and FA-2 are differenced on a data-split-by-data-split basis:

$$\Delta\text{RMSEV}^i = \text{RMSEV}^i_{\text{FA-2}} - \text{RMSEV}^i_{\text{LMC}}. \tag{24}$$

Here, the superscripted i corresponds to the ith data split. This same differencing is applied to the 300 first quartile and the 300 median RMSEV values.

Local mean centering outperforms (underperforms) FA-2 when $\Delta\text{RMSEV}^i > 0 (< 0)$. To answer question 1, we will simply count how many times the events $\Delta\text{RMSEV}^i > 0$ and $\Delta\text{RMSEV}^i < 0$ occur (these frequencies will be denoted by n_{OUT} and n_{UND}, respectively). To answer question 2, we will compute the average difference in performance. For $\Delta\text{RMSEV}^i > 0$ and $\Delta\text{RMSEV}^i < 0$, we respectively compute

$$\Delta_{\text{OUT}} = \frac{1}{n_{\text{OUT}}} \sum_{i=1}^{n_{\text{OUT}}} \Delta\text{RMSEV}^i \quad \text{and} \quad \Delta_{\text{UND}} = \frac{1}{n_{\text{UND}}} \sum_{i=1}^{n_{\text{UND}}} \Delta\text{RMSEV}^i.$$

Figure 7 displays n_{OUT} and Δ_{OUT} as the length and height of a white rectangle—see Figure 7. The same applies to n_{UND} and Δ_{UND} but with respect to a black rectangle. (The black rectangle lies beneath the x-axis since Δ_{UND} is negative.) The white and black rectangles associated with the pairs $\{n_{\text{OUT}}, \Delta_{\text{OUT}}\}$ and $\{n_{\text{UND}}, \Delta_{\text{UND}}\}$ are computed for each data set and

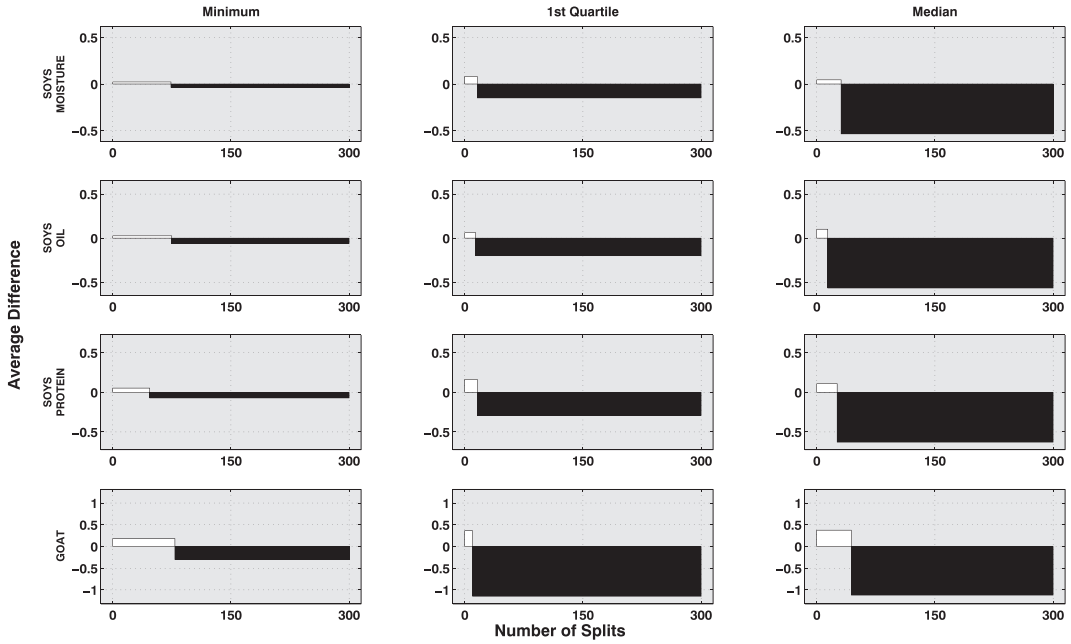

FIGURE 7 Comparison trend plots of between FA-2 and LMC. Average difference are plotted on *y*-axis with negative values indicating that LMC underperforms FA-2

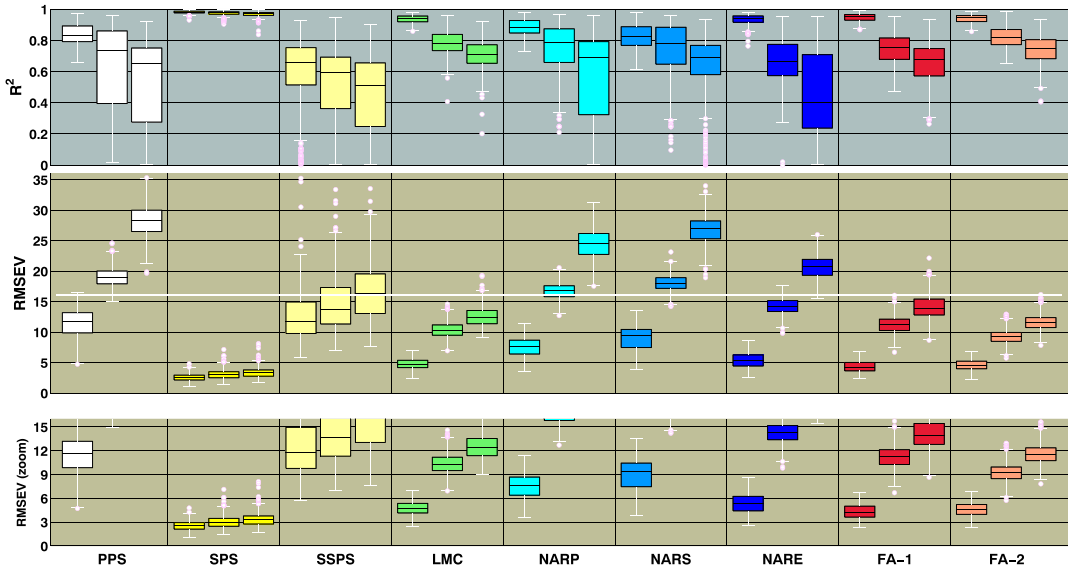

FIGURE 8 Performance boxplots for **GOAT** across calibration updating methods. Here, 10% of the secondary samples were used for the CALS sample set. See section 6.1 and Figure 1 for the description

for each quartile case (minimum, first quartile, and median). The wider the black rectangle is (wider than 150 splits) and the taller the black rectangle is, the more FA-2 outperforms LMC. For the GOAT and SOYS datasets, FA-2 outperforms LMC, on average, across all situations. Feature augmentation provides consistent improvement.

6.2.2 | GOAT

Figures 5, 6, 8, and 9 display the performance boxplots, median, and minimum RMSEV heatmaps, respectively, for the fecal goat data set. The GOAT data set shares many of the performance traits with the SOYS data set. As with SOYS, the FA collection performs the best overall—with FA-2 providing lower RMSEV and higher R^2 values; FA variant. Similarly, the lower quartile of the RMSEV values for NARE are comparable to the benchmark performance of LMC. The trends

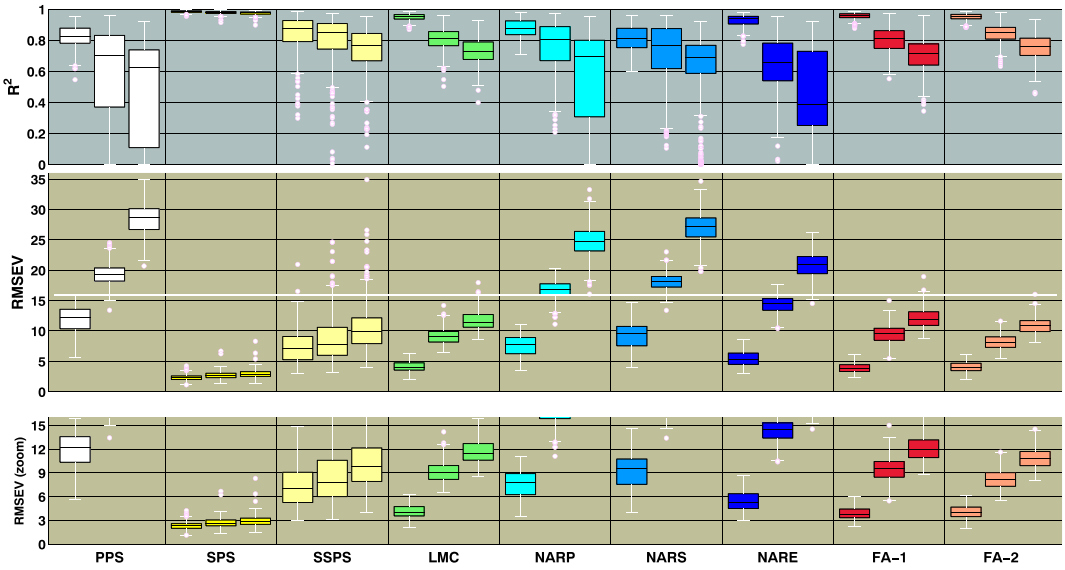

FIGURE 9 The performance boxplots for **GOAT** across calibration updating methods. Here, 20% of the secondary samples were used for the CALS sample set. See section 6.1 and Figure 1 for the description

in the performance comparison in Figure 7 for SOYS are similar for GOAT for LMC, FA-2, and NARE. Increasing the size CALS sample set $\{\mathbf{X}_S, \mathbf{y}_S\}$ from 10% to 20% marginally improves performance for some methods, but no meaningful improvement is observed for most methods.

7 | CONCLUSION AND FUTURE WORK

In this paper, we have created many calibration updating methods that are applicable to a wide range of data sets, especially those that do not contain standardization samples. Out of all the new calibration updating methods utilizing the CALS sample set $\{\mathbf{X}_S, \mathbf{y}_S\}$, the FA-2 variant performs either similar to LMC or provides improvement. In the NAR collection that exclusively use unlabeled secondary spectra, the NARE variant provides best results and is also parsimonious: the vertically augmented matrix $\mathbf{R}$ only consists of one row. (Although the NARE $\mathbf{R} = (\mu_P - \mu_{SU})^T$ matrix consists of only one row, the centroid μ_{SU} does use all available unlabeled secondary samples when constructing the mean.) This NAR mechanism of calibration updating is simple and powerful. In particular, the full utility of NAR cannot be fully addressed until many more unlabeled secondary samples are used such that the sample size greatly exceeds the number of labeled secondary samples (ie, $n_{SU} \gg n_S$). While NARE with limited unlabeled samples does not perform as well as FA-2 at the first quartile and median boxplots, this observation is based only on unlabeled data from the secondary samples. The user needs to assess the trade-off for improving accuracy against spending the time and resources to obtain the corresponding reference values. Additionally, future work will involve looking at NARE and the other NAR approaches in the transductive mode for improved models.

The other NAR variants, NARP and NARS, were only studied in the orthogonal projection mode, and using oblique projections may provide improved models. However, this would require an additional tuning parameter. For example, the oblique projection $\mathbf{R} = \mathbf{X}(I - \gamma \mathbf{V}_{SU}\mathbf{V}_{SU}^T)$ would require a determination of the tuning parameter γ. The current work with NARP and NARS described in this paper uses the default value of $\gamma = 1$.

From the heatmaps shown in Figures 2 and 3, it is clear that model selection is not easy. Specifically, numerous models exists that predict well including overfitted models at the higher PLS LVs. Ongoing work in our laboratory involves developing processes for model selection robust to data sets and updating method. Preliminary work has shown that fusion of multiple measures of model quality has potential.[48,57]

There are many avenues of exploration in this framework. In particular, we seek a more effective coupling of labeled and unlabeled secondary samples that improve performance beyond what can be achieved with labeled secondary samples alone. To date, no studies have been performed characterizing the degree of differences between primary and secondary before calibration will fail. Such work is also ongoing.

ACKNOWLEDGMENTS

Financial support was provided by the National Science Foundation Chemical Measurement and Imaging grant CHE-1506417. We kindly acknowledge John Walker at the Texas A&M AgriLife Research and Extension Center at San Angelo, TX, for making available the fecal goat data set for us to analyze.

ORCID

Erik Andries http://orcid.org/0000-0002-6962-2321
John H. Kalivas http://orcid.org/0000-0001-7056-976X

REFERENCES

1. Osborne BG, Fearn T. Collaborative evaluation of universal calibrations for the measurement of protein and moisture in flour by near-infrared reflectance. *Int J Food Sci Technol*. 1983;18(4):453-460.
2. Osborne BG, Fearn T. Collaborative evaluation of near-infrared reflectance analysis for the determination of protein, moisture and hardness in wheat. *J Sci Food Agric*. 1983;34(9):1011-1017.
3. Shenk JS, Westerhaus MO, Templeton WC. Calibration transfer between near-infrared reflectance spectrophotometers. *Crop Sci*. 1985;25(1):159-161.
4. Wang Y, Veltkamp DJ, Kowalski BR. Multivariate instrument standardization. *Anal Chem*. 1991;63(23):2750-2756.
5. Wang Y, Kowalski BR. Temperature-compensating calibration transfer for near-infrared filter instruments. *Anal Chem*. 1993;65(9):1301-1303.
6. de Noord OE. Multivariate calibration standardization. *Chemom Intell Lab Syst*. 1994;25(2):85-97.
7. de Noord OE. The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. *Chemom Intell Lab Syst*. 1994;23(1):65-70.
8. Bouveresse E, Massart DL. Improvement of the piecewise direct standardization procedure for the transfer of NIR spectra for multivariate calibration. *Chemom Intell Lab Syst*. 1996;32(2):201-213.
9. Walczak B, Bouveresse E, Massart DL. Standardization of near-infrared spectra in the wavelet domain. *Chemom Intell Lab Syst*. 1997;36(1):41-51.
10. Sjöblom J, Svensson O, Josefson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom Intell Lab Syst*. 1998;44(1-2):229-244.
11. Anderson CE, Kalivas JH. Fundamentals of calibration transfer through Procrustes analysis. *Appl Spectrosc*. 1999;53(10):1268-1276.
12. Feudale RN, Woody NA, Tan HW, Myles AJ, Brown SD, Ferré J. Transfer of multivariate calibration models: a review. *Chemom Intell Lab Syst*. 2002;64(2):181-192.
13. Fearn T. Standardisation and calibration transfer for near infrared instruments: a review. *J Near Infrared Spectrosc*. 2001;9(4):229-244.
14. Kalivas JH, Siano GG, Andries E, Goicoechea HC. Calibration maintenance and transfer using Tikhonov regularization approaches. *Appl Spectrosc*. 2009;63(7):800-809.
15. Kunz MR, Kalivas JH, Andries E. Model updating for spectral calibration maintenance and transfer using 1-norm variants of Tikhonov regularization. *Anal Chem*. 2010;82(9):3642-3649.
16. Sulub Y, Small GW. Spectral simulation methodology for calibration transfer of near-infrared spectra. *Appl Spectrosc*. 2007;61(4):406-413.
17. Haaland DM, Melgaard DK. New Prediction-Augmented Classical Least Squares (PACLS) methods: application to unmodeled interferents. *Appl Spectrosc*. 2000;54(9):1303-1312.
18. Igne B, Hurburgh CR. Standardisation of near infrared spectrometers: evaluation of some common techniques for intra- and inter-brand calibration transfer. *J Near Infrared Spectrosc*. 2008;16(6):539-550.
19. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. *Chemom Intell Lab Syst*. 1998;44(1-2):175-185.
20. Tan HW, Brown SD. Wavelet analysis applied to removing non-constant, varying spectroscopic background in multivariate calibration. *J Chemom*. 2002;16(5):228-240.
21. 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.
22. Igne B, Roger J-M, Roussel S, Bellon-Maurel V, Hurburgh CR. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemom Intell Lab Syst*. 2009;99(1):57-65.
23. Sales F, Callao MP, Ruis FX. Multivariate standardization techniques using UV-Vis data. *Chemom Intell Lab Syst*. 2002;38(1):63-73.
24. Swierenga H, Haanstra WG, de Weijer AP, Buydens LMC. Comparison of two different approaches toward model transferability in NIR spectroscopy. *Appl Spectrosc*. 1998;52(1):7-16.
25. Kalivas JH. Learning from Procrustes analysis to improve multivariate calibration. *J Chemom*. 2008;22:227-234.
26. Stork CL, Kowalski BR. Weighting schemes for updating regression models—a theoretical approach. *Chemom Intell Lab Syst*. 1999;48(2):151-166.

27. Brown CD. Discordance between net analyte signal theory and practical multivariate calibration. *J Chemom*. 2004;76(15):4364-4373.
28. Helland K, Berntsen HE, Borgen OS, Martens H. Recursive algorithm for partial least squares regression. *Chemom Intell Lab Syst*. 1992;14:129-137.
29. Hansen P-C. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA: SIAM Press; 1998.
30. Lawson CL, Hanson RJ. *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice Hall Press; 1974.
31. Ottaway J, Farrell JA, Kalivas JH. Spectral multivariate calibration without laboratory prepared or determined reference analyte values. *Anal Chem*. 2013;85:1509-1516.
32. Haaland DM, Melgaard DK. New augmented classical least squares for improved quantitative spectral analyses. *Vib Spectro*. 2002;29:171-175.
33. Haaland DM, Melgaard DK. New classical least squares/partial least squares hybrid algorithm for spectral analyses. *Appl Spectrosc*. 2001;55(1):1-8.
34. Melgaard DK, Haaland DM, Wehlburg CM. Concentration residual augmented classical least squares (CRACLS): a multivariate calibration method with advantages over partial least squares. *Appl Spectrosc*. 2002;56(5):615-624.
35. Wehlburg CM, Haaland DM, Melgaard DK, Martin LE. New hybrid algorithm for maintaining multivariate quantitative calibrations of a near-infrared spectrometer. *Appl Spectrosc*. 2002;56(5):605-614.
36. Westerhaus MO. Improving repeatability of NIR calibrations across instruments. In: Biston R, Bartiaux-Thill, eds. *Proceedings of the Third International Near Infrared Spectroscopy Conference*. Gembloux, Belgium: Agriculture Research Centre Publishing; 1991:671-674.
37. Andries E, Kalivas JH. Interrelationships between generalized Tikhonov regularization, generalized net analyte signal, and generalized least squares for desensitizing a multivariate calibration to interferences. *J Chemom*. 2013;27(5):249-260.
38. Kalivas JH, Brownfield B, Karki B. Sample-wise spectral multivariate calibration desensitized to new artifacts relative to the calibration data using a residual penalty. *J Chemom*. 2017;31(4):e2873.
39. Chapelle O, Schökopf B, Zien A. *Semi-supervised Learning*. Cambridge, MA: The MIT Press; 2006.
40. Thomas EV. Incorporating auxiliary predictor variation in principal component regression models. *J Chemom*. 1995;9(6):471-481.
41. Koren Y. Robust linear dimensionality reduction. *IEEE Trans Visual Comput Graphics*. 2004;10(4):459-470.
42. Ji S, Ye J. Generalized linear discriminant analysis: a unified framework and efficient model selection. *IEEE Trans Neural Networks*. 2008;19(10):1768-1782.
43. van der Maaten LJP, Postma EO, van den Herik HJ. Dimensionality reduction: a comparative review. *J Mach Learn Res*. 2009;10(1):66-71.
44. Longsine DE, McCormick S. Simultaneous Rayleigh-quotient minimization methods for $\mathbf{Ax} = \lambda\mathbf{Bx}$. *Linear Algebra Appl*. 1980;34:195-234.
45. Sameh AH, Wisniewski JA. A trace minimization algorithm for the generalized eigenvalue problem. *SIAM J Numer Anal*. 1982;19(6):1243-1259.
46. Andries E. Penalized eigendecompositions: motivations from domain adaptation for calibration transfer. *J Chemom*. 2017;31(4):e2818(1-14).
47. Ghifary M. Domain adaptation and domain generalization with representation learning. *Ph.D. Thesis*: Victoria University of Wellingtion; 2016.
48. Tencate AJ, Kalivas JH, Andries E. Penalty processes for combining roughness and smoothness in spectral multivariate calibration. *J Chemom*. 2016;30(4):144-152.
49. Chen A, Owen AB, Shi M. Data enriched linear regression. *Electron J Statist*. 2015;9(1):1078-1112.
50. Chung J, Español MI, Nguyen T. Optimal regularization parameters for general-form Tikhonov regularization. arXiv:1407.1911; 2014.
51. Whitworth WR. Estimating redberry and ashe juniper intake using fecal NIRS. *Master Thesis*: Angelo State University; 2002.
52. Walker JW, Campbell ES, Lupton CJ, Taylor CA, Waldron DF, Landau SY. Effects of breed, sex and age on the variation and ability of fecal near-infrared reflectance spectra to predict the composition of goat diets. *J Anim Sci*. 2007;85(2):518-526.
53. Forina M, Drava G, Armanino C, et al. Transfer of calibration function in near-infrared spectroscopy. *Chemom Intell Lab Syst*. 1995;27(2):189-203.
54. Bouveresse E, Hartmann C, Massart DL. Standardization of near-infrared spectrometric instruments. *Anal Chem*. 1995;68(6):982-990.
55. Vitale R, Westerhauis JA, Næs T, Smilde AK, de Noord OE, Ferrer A. Selecting the number of factors in principal component analysis by permutation testing—numerical and practical aspects. *J Chemom*. 2017;31:e2937.
56. Malinowski ER. Determination of rank by augmentation (DRAUG). *J Chemom*. 2010;25:323-328.
57. Tencate AJ, Kalivas JH, White AJ. Fusion strategies for selecting multiple tuning parameters for multivariate calibration and other penalty based processes: a model updating application for pharmaceutical analysis. *Anal Chim Acta*. 2017;921:28-37.

How to cite this article: Andries E, Kalivas JH, Gurung A. Sample and feature augmentation strategies for calibration updating. *Journal of Chemometrics*. 2018;e3080. https://doi.org/10.1002/cem.3080