Research Article

CHEMOMETRICS

Received: 1 November 2012, Revised: 1 March 2013, Accepted: 22 March 2013, Published online in Wiley Online Library: 29 April 2013

(wileyonlinelibrary.com) DOI: 10.1002/cem.2501

Interrelationships between generalized Tikhonov regularization, generalized net analyte signal, and generalized least squares for desensitizing a multivariate calibration to interferences

Erik Andries[a,b*] and John H. Kalivas[c]

Orthogonal pre-processing (orthogonal projection) of spectral data is a common approach to generate analyte-specific information for use in multivariate calibration. The goal of this pre-processing is to remove from each spectrum the respective sample interferent contributions (spectral interferences from overlap, scatter, noise, etc.). Two approaches to accomplish orthogonal pre-processing are net analyte signal (NAS) and generalized least squares (GLS). Developed in this paper is the mathematical relationship between NAS and GLS. It is also realized that orthogonal NAS pre-processing can remove too much analyte signal and that the degree of interferent correction can be regulated. Similar to GLS, the degree of correction is accomplished by using a regularization (tuning) parameter to form generalized NAS (GNAS). Also developed in this paper is an alternative to GNAS and GLS based on generalized Tikhonov regularization (GTR). The mathematical relationships between GTR, GNAS, and GLS are derived. A result is the ability to express the model vector as the sum of two contributions: the orthogonal NAS contribution and a non-NAS contribution from the interferent components. Thus, rather than the usual situation of sequentially pre-processing data by either GNAS or GLS followed by model building with the pre-processed data, the methods of GTR, GNAS, and GLS are expressed as direct computations of model vectors allowing concurrent pre-processing and model building to occur. Simultaneous pre-processing and model forming are shown to be natural to the GTR process. Two near-infrared spectroscopic data sets are studied to compare the theoretical relationships between GTR, GNAS, and GLS. One data set covers basic calibration, and the other data set is for calibration maintenance. Filter factor representation is key to developing the interprocess relationships.

Keywords: multivariate calibration; pre-processing; net analyte signal; NAS; orthogonal pre-processing; generalized least squares; GLS; Tikhonov regularization; TR; filter factors

1. INTRODUCTION

Multivariate calibration relates a dependent variable such as a chemical or physical property to independent variables such as spectroscopic measurements via the model vector [1–3]. The model vector is commonly estimated by the methods of partial least squares (PLS), the Tikhonov regularization (TR) variant of ridge regression (RR), or principal component regression (PCR). The goal in spectral calibration is to determine an appropriate estimate of the model vector in order to balance figure of merits such as prediction accuracy and precision (the bias/variance trade-off), sensitivity, and detection limit. This balance is typically accomplished by forming the model vector such that it is desensitized to the interferent space [4,5]. The interferent space is that part of the calibration space not due to the analyte and can consist of any combination of effects from chemical (including spectral), physical, environmental, or instrumental sources.

To assist in developing a model insensitive to the interferent space, pre-processing methods are commonly used to remove the interferent information [6,7]. The methods are varied and well reviewed including examples and lists of pros and cons [7].

* *Correspondence to: E. Andries, Department of Mathematics, Central New Mexico Community College, Albuquerque, NM 87106, USA*
Email: erik.andries@gmail.com

a *E. Andries*
Department of Mathematics, Central New Mexico Community College, Albuquerque, NM 87106, USA

b *E. Andries*
Center for Advanced Research Computing, University of New Mexico, Albuquerque, NM 87106, USA

c *J. H. Kalivas*
Department of Chemistry, Idaho State University, Pocatello, ID 83209, USA

The methods can be as simple as using derivatives based on only spectral information or more complicated such as orthogonal signal correction that use spectral and analyte information—see [7] and references therein.

Of the many pre-processing possibilities, two common approaches are net analyte signal (NAS) processing and generalized least squares (GLS) [7–9]. In both cases, the goal is to remove from spectra the interferent contributions (see Section 4 for the respective details). Essential to both processes is a set of spectra spanning the interferent sample matrix, that is, spectra characterizing artifacts that the model needs to be desensitized to. Examples of orthogonal NAS pre-processing applications include instrumentation differences [10,11], replicate measurement variation such as in probe placement [12], baseline variations [13,14], temperature [15,16], particle scattering [17], optical path-length correction [18], and human subject differences in biomedical studies [19–21].

When spectra are pre-processed, it is generally found that fewer reference samples are needed to form the calibration model [9,22], and this has also been found to be true for a TR variant in which no reference samples are used [23]. Pre-processing also often results in simpler models, for example, fewer PLS latent variables, but prediction errors are commonly the same or slightly smaller than from models formed without pre-processing. With simpler models, it is typically indicated that improved model interpretation is possible.

Whereas GLS provides flexibility in the degree of pre-processing spectra relative to the interferent space, NAS has only been used as an orthogonal pre-processor. As noted in [4], orthogonal correction to spectra is not always best for reducing prediction errors in the inverse least squares setting such as with PLS or RR. Complete orthogonal pre-processing is useful when there is little interferent contribution to the measured spectra. Instead, it is suggested to include the correction spectra (interferent spectra) in the modeling process as with the TR variants [4,5,11,24] recently reviewed [25] or other approaches [26,27].

Some limited interrelationships between TR and NAS as well as GLS and NAS have been presented in spectral pre-processing formats [9,10]. In both cases, the relationships are derived and discussed only at the point where TR or GLS converge to NAS pre-processing. In [10], a generalized form of NAS (GNAS) was presented (see Section 4.1), but only orthogonal NAS was studied. Henceforth, NAS shall refer to the orthogonal version as detailed in Section 4.1.

Developed in this paper are generalized TR (GTR) and detailed mathematical derivations of the interrelationships between GTR, GNAS, and GLS. Interrelationships are established in two formats—*spectral processing* and *inverse processing*. Spectral processing considers GNAS, GLS, and GTR as spectral pre-processing methods that replace the linear system $\mathbf{Xb} = \mathbf{y}$ with a modified linear system $\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}}$ that incorporates interferent space information. The modified linear system is then solved by conventional means (PLS, TR, or PCR), and a model vector $\mathbf{b}$ is obtained. Inverse processing works with GNAS, GLS, and GTR as model-building methods that essentially pre-process the data while forming the model. The spectral and inverse processing methods are applied to two spectral data sets demonstrating equivalencies as well as respective distinctiveness. The advantages and disadvantages of the three approaches are described. One spectral data set is concerned with spectral interferences primarily from spectral overlap, and the other spectral set deals with instrument interferences, that is, differences between instruments are used to form a new model robust to the instrument differences.

2. PRELIMINARIES

The paper's focus is theoretical development of interrelationships between GNAS, GLS, and GTR. Evaluation of the two data sets is not to determine whether GNAS, GLS, or GTR is better than no pre-processing or to ascertain if the approaches form more interpretable models compared with no pre-processing. Instead, the data sets are used to evaluate the theoretical relationships. Comparisons of the processes are based on prediction errors. Other figures of merit could be evaluated, for example, sensitivity and limit of detection, but such merits were not evaluated.

2.1. Notation

Lowercase and uppercase symbols that are not boldface represent scalars (x or P). Lowercase and uppercase boldface symbols represent column vectors ($\mathbf{x}$) and matrices ($\mathbf{X}$), respectively. The superscripted symbols T, −1, and + indicate the transpose, inverse, and pseudo-inverse, respectively, of a vector or matrix. A vector of n ones or zeros is indicated by $\mathbf{1}_n$ and $\mathbf{0}_n$, respectively, whereas $\mathbf{I}_n$ represents the identity matrix of dimension n. An $m \times p$ matrix $\mathbf{A}$ can be formed by concatenating its p column vectors $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p]^{\mathrm{T}}$, whereas a diagonal matrix is indicated via the "diag" notation, for example, $\mathbf{I}_n = \mathrm{diag}(\mathbf{1}_n) = \mathrm{diag}(1, 1, \ldots, 1)$. A zero matrix of dimension $p \times q$ will be denoted by $\mathbf{0}_{p,q}$ (if $p \times 1$, then $\mathbf{0}_p$). The element associated with the ith row and jth column of the matrix $\mathbf{A}$ will be denoted in two ways: a_{ij} or $(A)_{ij}$. In this paper, the $m \times n$ matrix $\mathbf{X}$ represents calibration data of n absorbance measurements across m samples such that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]^{\mathrm{T}}$, where $\mathbf{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^{\mathrm{T}}$. The vector $\mathbf{y} = [y_1, y_2, \ldots, y_m]^{\mathrm{T}}$ represents the response variables (e.g., analyte concentrations) across samples. The $k \times n$ matrix $\mathbf{L}$ is a matrix of interferent spectra—spectra not due to analyte variation but only from spectral interference. These spectra can result from any combination of effects ranging from chemical, physical, spectral, environmental, to instrumental sources. If $\mathbf{X}$ is from one set of measurement conditions (primary conditions), then $\mathbf{L}$ can be from the primary conditions for primary calibration or $\mathbf{L}$ can be from the new measurement conditions (secondary conditions) for calibration maintenance. Before discussing how spectra can be processed, it will be convenient to discuss the standard matrix decompositions associated with $\mathbf{L}$.

2.2. Singular value decomposition of interferent spectra

Let $\mathbf{L}$ be a matrix of interferent spectra of dimension $k \times n$ ($k << n$). The *full* singular value decomposition (SVD) of $\mathbf{L}$ is defined as

$$\mathbf{L} = \mathbf{H}\hat{\mathbf{R}}\hat{\mathbf{Q}}^{\mathrm{T}}, \quad \text{where} \quad \hat{\mathbf{R}} = [\mathbf{R}, \mathbf{0}_{k,q}] \ (q = n - k), \quad \hat{\mathbf{Q}} = [\mathbf{Q}, \mathbf{Q}_{\perp}] \qquad (1)$$

The matrices $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_k]$ and $\hat{\mathbf{Q}} = [\mathbf{q}_1, \ldots, \mathbf{q}_n]$ have dimensions $k \times k$ and $n \times n$, respectively, and are orthonormal. The diagonal matrix $\mathbf{R} = \mathrm{diag}(r_1, \ldots, r_k)$ contains the singular values of $\mathbf{L}$. The matrix $\hat{\mathbf{Q}}$, the columns of which are the loading vectors of $\mathbf{L}$, can be split such that $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_k]$ forms an orthonormal basis for $\mathcal{L}$, the space spanned by the spectral samples in $\mathbf{L}$, whereas $\mathbf{Q}_{\perp} = [\mathbf{q}_{k+1}, \ldots, \mathbf{q}_n]$ forms an orthonormal basis for $\mathcal{L}_{\perp}$, the space spanned by the spectral samples orthogonal to $\mathcal{L}$.

The matrix $\mathbf{L}$ is often written more compactly in terms of the reduced SVD such that $\mathbf{L} = \mathbf{H}\hat{\mathbf{R}}\hat{\mathbf{Q}}^{\mathrm{T}} = \mathbf{H}\mathbf{R}\mathbf{Q}^{\mathrm{T}}$.

3. SPECTRAL PROCESSING

As noted in Section 1, the spectral processing approach replaces the linear system $\mathbf{X}\mathbf{b} = \mathbf{y}$ with the modified linear system $\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}}$ that incorporates interferent space information. The modification of the linear system can occur in one of two ways—by post-multiplication or augmentation. Post-multiplication, as its name implies, post-multiplies $\mathbf{X}$ by a matrix $\mathbf{P}$ involving the matrix $\mathbf{L}$ of interferent spectra such that the modified linear system takes the form

$$\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}} \quad \text{or} \quad (\mathbf{X}\mathbf{P})\mathbf{b} = \mathbf{y}, \quad \text{where} \quad \widetilde{\mathbf{X}} = \mathbf{X}\mathbf{P} \quad \text{and} \quad \widetilde{\mathbf{y}} = \mathbf{y} \tag{2}$$

Augmentation instead solves the linear system

$$\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}} \quad \text{or} \quad \begin{bmatrix} \mathbf{X} \\ \mathbf{P} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_k \end{bmatrix}, \quad \text{where} \quad \widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{P} \end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_k \end{bmatrix}$$

We will first examine the post-multiplication approach.

3.1. Spectral processing via post-multiplication

For the post-multiplication approach, a conventional multivariate calibration technique (e.g., PLS, TR, or PCR) solves the modified linear system $\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}}$ or, equivalently, the linear system $(\mathbf{X}\mathbf{P})\mathbf{b} = \mathbf{y}$. Different processing matrices $\mathbf{P}$ yield different approaches. In this paper, we examine the following:

$$\mathbf{P} = \begin{cases} \mathbf{I}_n - \mathbf{L}^{\dagger}\mathbf{L}, & \text{NAS} \\ \mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L}, & \text{GNAS} \\ \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)^{-1/2}, & \text{GLS} \end{cases} \tag{3}$$

We will see in subsequent sections that all three methods share a similar processing mechanism whereby each loading vector $\mathbf{q}_i$ of $\mathbf{L}$ is weighted in some fashion:

$$\mathbf{P} = \mathbf{I}_n - \mathbf{Q}\mathbf{F}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}_n - \sum_{i=1}^{k} f_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}, \quad \mathbf{F} = \mathrm{diag}(f_1, \ldots, f_k) \tag{4}$$

Different diagonal entries for f_i generate the NAS, GNAS, or GLS processing schemes.

3.1.1. *Net analyte signal spectral processing*

Net analyte signal processing is the simplest and most common pre-processing scheme. In short, one decomposes the calibration spectra into two orthogonal pieces, $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 = \mathbf{X}(\mathbf{Q}\mathbf{Q}^{\mathrm{T}}) + \mathbf{X}\left(\mathbf{I}_n - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)$$

where $\mathbf{X}_1 = \mathbf{X}(\mathbf{Q}\mathbf{Q}^{\mathrm{T}})$ and $\mathbf{X}_2 = \mathbf{X}\left(\mathbf{I}_n - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right)$ belong to $\mathcal{L}$ and $\mathcal{L}_{\perp}$, respectively. Instead of using $\mathbf{X}$, one throws away the interferent component $\mathbf{X}_1$ and keeps only the "interferent-free" spectra $\mathbf{X}_2$. One then solves the following linear system:

$$\mathbf{X}_{\mathrm{NAS}}\mathbf{b} = \mathbf{y}, \quad \text{where} \quad \mathbf{X}_{\mathrm{NAS}} = \mathbf{X}_2 = \mathbf{X}\left(\mathbf{I}_n - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right) \tag{5}$$

By making the processed spectra orthogonal to the space spanned by the interferent noise, one hopes to immunize the resulting model vector $\mathbf{b}$ against chemical or physical noise. It should be noted that Equation (5) uses all of the k vectors $\mathbf{q}_i$ of $\mathbf{Q}$. However, a subset of the available k vectors can be used instead, and this subset is often used in practice when the number of spectra in $\mathbf{L}$ is large.

Under the diagonal weighting scheme of Equation (4), the processing matrix $\mathbf{P}$ in Equation (5) gives unit weight to all of the loading vectors:

$$\mathbf{P} = \mathbf{I}_n - \mathbf{Q}\mathbf{Q}^{\mathrm{T}} = \mathbf{I}_n - \sum_{i=1}^{k} f_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}, \quad \text{where} \quad f_i = 1 \text{ for all } i$$

However, completely removing all of the interferent information from $\mathbf{X}$ can have deleterious effects on the calibration [4]. If the interferent spectra $\mathbf{L}$ strongly overlap with the pure-component spectrum associated with the analyte of interest in $\mathbf{X}$, then the NAS projection of $\mathbf{X}$ can substantially remove the signal that we want to capture. Hence, calibration with interferent spectra involves a fundamental trade-off: How much of the interferent and/or analyte information do we want to remove via a projection?

3.1.2. *Generalized net analyte signal spectral processing*

If one wants to process spectra that do not entirely eliminate interferent information so as to preserve some analyte signal, then one might opt for the GNAS processing treatment

$$\mathbf{X}_{\mathrm{GNAS}}\mathbf{b} = \mathbf{y}, \quad \text{where} \quad \mathbf{X}_{\mathrm{GNAS}} = \mathbf{X}\mathbf{P}, \quad \mathbf{P} = \mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L} \quad \text{with} \quad \beta \neq 1 \tag{6}$$

Under the diagonal weighting scheme of Equation (4), the GNAS processing matrix $\mathbf{P}$ of Equation (6) weights each loading vector equally but without unit weight:

$$\mathbf{P} = \mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L} = \mathbf{I}_n - \sum_{i=1}^{k} f_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}, \quad \text{where} \quad f_i = \beta \text{ for all } i \tag{7}$$

The value of β has occurred within the interval $[0, 1]$ as 0 indicates no removal of interferent information and 1 indicates the orthogonal removal of interferent information [9]. However, as shown in Section 5, the mathematical range of β can occur outside this interval as well. Moreover, the processed spectra $\mathbf{X}_{\mathrm{GNAS}}$ asymptotically approaches the NAS-processed spectra $\mathbf{X}_{\mathrm{NAS}}$ as β approaches 1.

3.1.3. *Generalized least squares spectral processing*

The GLS method uses the covariance information of $\mathbf{L}$ to process the spectra $\mathbf{X}$ (see Appendix A for a brief description of the mathematical derivation of the GLS processing matrix). The modified linear system for GLS is as follows:

$$\mathbf{X}_{\mathrm{GLS}}\mathbf{b} = \mathbf{y}, \quad \text{where} \quad \mathbf{X}_{\mathrm{GLS}} = \mathbf{X}\mathbf{P}, \quad \mathbf{P} = \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)^{-1/2} \tag{8}$$

It is not easy to understand, by just looking at Equation (8), how the matrix $\mathbf{P}$ precisely processes the spectra $\mathbf{X}$. To gain some

insight, we again call upon the SVD of **L** and the GLS inverse formula in Appendix C:

$$\mathbf{P} = (\mathbf{I}_n + \gamma \mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1/2} = \mathbf{I}_n - \sum_{i=1}^{k} f_i \mathbf{q}_i \mathbf{q}_i^{\mathrm{T}}, \quad f_i = 1 - \frac{1}{\sqrt{1+\gamma r_i^2}} \tag{9}$$

Unlike NAS and GNAS, GLS weights the loading vectors of **L** in a disproportionate manner. As the singular values r_i are ordered in descending order ($r_1 \geq \cdots \geq r_k$), the diagonal elements of $\mathbf{F} = \mathrm{diag}(f_1, \ldots, f_k)$ are similarly ordered such that $f_1 \geq \cdots \geq f_k$. As a result, GLS gives greater weight to the first few loading vectors of **L** and damps the remaining loading vectors that explain the least amount of variance. Again, as with NAS and GNAS, the k loading vectors of **Q** can be replaced with a subset of the loading vectors $\mathbf{q}_i$.

Although $f_1 \geq \cdots \geq f_k$, all of the diagonal elements f_i rapidly approach 1 when γ becomes sufficiently large, that is, $\mathbf{X}_{\mathrm{GLS}}$ approaches $\mathbf{X}_{\mathrm{NAS}}$ when γ goes to ∞. Note that Equation (9) only makes sense when $1 + \gamma r_i^2 > 0$ for all i. Hence, γ is effectively bounded in the interval $(-r_1^{-2}, \infty)$. When γ is non-negative, the diagonal weighting elements f_i are positive and bounded within [0, 1]. When γ is negative $(-1/r_1^2 < \gamma < 0)$, f_i is negative and bounded within the interval $(-\infty, 0)$.

Given that GLS and GNAS share a diagonal weighting mechanism, one may wonder if there is a relationship between β in Equation (7) and γ in Equation (9). The authors in [9] state that GNAS and GLS are equivalent when

$$\gamma = \left(\frac{1}{1-\beta}\right)^2 - 1 \tag{10}$$

However, this equation needs further elaboration. If GNAS is equivalent to GLS when Equation (10) is true, then *all* of the corresponding weights f_i in Equations (7) and (9) would have to be equal as well, that is,

$$f_i = \beta = 1 - \frac{1}{\sqrt{1+\gamma r_i^2}} \quad \text{for all } i \tag{11}$$

As β is constant, the only way for equality to hold in Equation (11) is for all of the singular values r_i to be the same as well—an unlikely happenstance. One way for equality to hold is to artificially force all of the singular values r_i to be the same. For example, if one replaces **L** with its orthogonal loading vectors $\mathbf{Q}^{\mathrm{T}}$ (i.e., $\mathbf{L} := \mathbf{Q}^{\mathrm{T}}$), then the SVD of **L** would yield $\mathbf{R} = \mathbf{I}_k$ and

$$\beta = 1 - \frac{1}{\sqrt{1+\gamma}} \tag{12}$$

Solving for γ in Equation (12) results in Equation (10).

3.2. Spectral processing via augmentation

In the augmentation approach, the linear system

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{P} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_n \end{bmatrix} \quad \text{with} \quad \mathbf{P} = \lambda \mathbf{L} \quad \text{and} \quad \lambda \geq 0 \tag{13}$$

is solved instead of $\mathbf{Xb} = \mathbf{y}$. We will refer to this augmented linear system as GTR. The solution vector **b** of the GTR formulation in Equation (13) is equivalent to the regression vector that minimizes

$$\min \phi_\lambda(\mathbf{b}), \quad \text{where} \quad \phi_\lambda(\mathbf{b}) = \|\mathbf{Xb} - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{Lb}\|_2^2 \tag{14}$$

The preceding minimization essentially approximates the following equality-constrained least squares problem: solve $\mathbf{Xb} = \mathbf{y}$ subject to the equality constraints $\mathbf{Lb} = \mathbf{0}_k$ [28]. Instead of trying to rigorously satisfy both linear systems to a high degree of accuracy, GTR seeks a trade-off between minimizing the residual norm $\|\mathbf{Xb} - \mathbf{y}\|_2$ and the equality constraint norm $\|\mathbf{Lb}\|_2$.

Geometrically, the constraint norm is a statement about orthogonality: as noted in [10] with $m \geq n$, the larger λ is, the greater emphasis we place on $\mathbf{Lb} = \mathbf{0}_k$, that is, the more we require **b** to be perpendicular to $\mathcal{L}$—the space spanned by the spectra in **L**. The same is true when $m < n$ except that the solution of Equation (13) requires additional regularization—through either PCR or PLS in which the number of latent vectors is the regularization parameter or through an additional L2-penalty or ridge parameter via TR [25]. The first approach, the solution by PLS, is used in this study.

Spectroscopically, the constraint $\mathbf{Lb} = \mathbf{0}$ attempts to immunize the model vector against noise by orthogonally pointing **b** away from the noise space spanned by spectral interferents (see [4] and references therein). When λ approaches infinity, the GTR solution approaches the NAS solution as the orthogonality condition $\mathbf{Lb} = \mathbf{0}$ is given greater and greater weight. This observation forms the basis of many calibration maintenance and transfer methods (see [24,29,30] and the references contained therein) and augmented classical least squares [31–35] procedures, which decompose spectra into pure-component concentrations and pure-component spectra. In summary, with GTR, concurrent processing occurs with modeling. As λ increases away from zero, the degree of orthogonal correction to **b** increases, that is, **b** becomes less oblique to $\mathcal{L}$.

4. INVERSE PROCESSING

Spectral processing replaces the primary spectra **X** with processed spectra in order to generate a calibration model that is hopefully resistant to chemical and physical interferents. However, the resulting model vector **b** obtained by these multivariate calibration techniques do not give much insight on how **b** depends on the interferent spectra **L** and the associated parameters β, γ, or λ. In this section, we explicitly compute closed-form expressions for GNAS, GLS, and GTR model vectors (termed $\mathbf{b}_{\mathrm{INV}}$) via pseudo-inverses of **X** and **L**. The construction of the model vector **b** will be referred to as *inverse processing*. Unlike spectral processing where the spectra were decomposed into two perpendicular components, inverse processing instead decomposes the model vector into two perpendicular terms

$$\mathbf{b}_{\mathrm{INV}} = \mathbf{b}_{\mathrm{REG}} + \mathbf{b}_{\mathrm{NAS}}, \quad \mathbf{b}_{\mathrm{REG}} \in \mathcal{L} \quad \text{and} \quad \mathbf{b}_{\mathrm{NAS}} \in \mathcal{L}_\perp \tag{15}$$

such that $\mathbf{b}_{\mathrm{INV}}$ denotes the solution obtained by inverse processing. The first term $\mathbf{b}_{\mathrm{REG}}$ is the only term that depends upon the processing-specific regularization parameters β, γ, or λ. The term regularization generally refers to the process of how these parameters shrink the size of the model vector to obtain a statistically or spectroscopically desired solution. The second term $\mathbf{b}_{\mathrm{NAS}}$ is in fact the NAS solution, and it is completely unaffected by the choice of these regularization parameters. In short, we show that the model

vectors obtained by GNAS, GLS, and GTR inverse processing are just nearby perturbations of the NAS solution.

Note that Equation (15) is true for other model vectors obtained without GNAS, GLS, or GTR processing, for example, using PLS with the original **X**. Thus, with Equation (15), it is possible to determine the degree of non-orthogonality (obliqueness) of the final vector to the interferents. To accomplish this analysis, the interferent spectra **L** would be needed to compute $\mathbf{b}_{NAS}$. Mathematically, if $\mathbf{b}_{PLS}$ is a PLS model vector computed using the original **X** without processing, then

$$\mathbf{b}_{REG} = \mathbf{b}_{PLS} - \mathbf{b}_{NAS}$$

and $\mathbf{b}_{REG}$ would be represent the degree that $\mathbf{b}_{PLS}$ is shifted from orthogonality defined by $\mathbf{b}_{NAS}$.

4.1. Inverse processing by pre-multiplication

In spectral processing, we post-multiply **X** by **P** and then apply a multivariate calibration procedure to the linear system $(\mathbf{XP})\mathbf{b} = \mathbf{y}$ to obtain **b**. In the inverse processing, the sequence of processing steps changes. When we take the pseudo-inverse of **XP** and solve for **b**,

$$(\mathbf{XP})\mathbf{b} = \mathbf{y} \quad \Rightarrow \quad \mathbf{b} = (\mathbf{XP})^{\dagger}\mathbf{y} = \mathbf{P}^{\dagger}\mathbf{g}, \quad \text{where} \quad \mathbf{g} = \mathbf{X}^{\dagger}\mathbf{y} \tag{16}$$

we are not processing the spectra **X** but instead the ordinary least squares solution **g** is processed by pre-multiplication with $\mathbf{P}^{\dagger}$.

The ordinary least squares solution $\mathbf{X}^{\dagger}\mathbf{y}$ in Equation (16) does not typically yield the best spectroscopically plausible solution. As a result, $\mathbf{g} = \mathbf{X}^{\dagger}\mathbf{y}$ can be replaced with a regularized least squares solution generated by PLS or PCR (or some other multivariate calibration technique). In this paper, the vector **g** will be replaced with a PLS model vector.

4.1.1. *Net analyte signal inverse processing*

The pseudo-inverse of the NAS processing matrix $\mathbf{P} = \mathbf{I}_n - \mathbf{QQ}^{\mathsf{T}}$ is itself. Hence, the NAS inverse processing solution in Equation (16) is

$$\mathbf{b}_{NAS} = \mathbf{P}^{\dagger}\mathbf{g} = (\mathbf{I}_n - \mathbf{QQ}^{\mathsf{T}})\mathbf{g} = \left(\mathbf{Q}_{\perp}\mathbf{Q}_{\perp}^{\mathsf{T}}\right)\mathbf{g} \tag{17}$$

From this perspective, NAS inverse processing projects the least squares solution **g** into $\mathcal{L}_{\perp}$—the subspace orthogonal to the subspace spanned by spectral interferents.

4.1.2. *Generalized net analyte signal inverse processing*

The solution **b** in Equation (16) using GNAS inverse processing is generically expressed as the following:

$$\mathbf{b}_{GNAS} = \mathbf{P}^{\dagger}\mathbf{g} = (\mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L})^{-1}\mathbf{g}, \quad \beta \neq 1 \tag{18}$$

Using the SVD of **L** and the inversion formula in Appendix C, the GNAS solution can be split into two solution components:

$$\mathbf{b}_{GNAS} = \mathbf{b}_{REG} + \mathbf{b}_{NAS} \tag{19}$$

where

$$\mathbf{b}_{REG} = \mathbf{QFQ}^{\mathsf{T}}\mathbf{g}, \quad \mathbf{F} = \frac{1}{1-\beta}\mathbf{I}_k = \text{diag}(f_1, \ldots, f_k), \quad \text{where} \quad f_i = \frac{1}{1-\beta} \tag{20}$$

and $\mathbf{b}_{NAS}$ is the same solution as in Equation (17). Note that the solution component $\mathbf{b}_{REG}$ is an *oblique* projection of **g** onto the space spanned by **L**. Recall that **g** in Equation (20) is defined in Section 4.1 as a PLS model vector.

The solution $\mathbf{b}_{GNAS}$ will approach $\mathbf{b}_{NAS}$ when $\mathbf{b}_{REG}$ goes to zero. Recall that with NAS-processed spectra, $\mathbf{X}_{GNAS} = \mathbf{X}(\mathbf{I}_n - \beta\mathbf{QQ}^{\mathsf{T}})$ approaches $\mathbf{X}_{NAS}$ as β approaches 1. However, the GNAS inverse solution does not share this limiting behavior. When β approaches 1, the regularized component $\mathbf{b}_{REG} = (1-\beta)^{-1}\mathbf{QQ}^{\mathsf{T}}\mathbf{g}$ "blows up" to $\pm\infty$ (a consequence of the inverse matrix $\mathbf{P}^{-1} = \left(\mathbf{I}_n - \beta\mathbf{QQ}^{\mathsf{T}}\right)^{-1}$ becoming increasingly rank deficient and singular) and $\mathbf{b}_{GNAS}$ diverges away from $\mathbf{b}_{NAS}$. For $\mathbf{b}_{REG}$ to converge to zero, $|\beta|$ must approach ∞. Hence, GNAS has an asymptotic dichotomy, depending upon whether **X** or **g** is being processed.

4.1.3. *Generalized least squares inverse processing*

The solution **b** in Equation (16) using GLS inverse processing is the following:

$$\mathbf{b}_{GLS} = \mathbf{P}^{\dagger}\mathbf{g} = \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathsf{T}}\mathbf{L}\right)^{1/2}\mathbf{g} \tag{21}$$

Using the SVD of **L** and the inversion formula in Appendix C, the GLS solution can also be expressed as the sum of two terms:

$$\mathbf{b}_{GLS} = \mathbf{b}_{REG} + \mathbf{b}_{NAS} \tag{22}$$

where $\mathbf{b}_{NAS}$ is the same solution as in Equation (17) and

$$\mathbf{b}_{REG} = \mathbf{QFQ}^{\mathsf{T}}\mathbf{g}, \text{ where } \mathbf{F} = \left(\mathbf{I}_k + \gamma\mathbf{R}^2\right)^{\frac{1}{2}} = \text{diag}(f_1, \ldots, f_k), \; f_i = \sqrt{1+\gamma r_i^2} \tag{23}$$

Again, we remind the reader that **g** in Equation (23) is defined in Section 4.1 as a PLS model vector.

As in the GNAS case, there is a similar asymptotic dichotomy with GLS as well. When the processed spectra $\mathbf{X}_{GLS} = \mathbf{X}(\mathbf{I}_n + \gamma\mathbf{L}^{\mathsf{T}}\mathbf{L})^{-1/2}$ approaches $\mathbf{X}_{NAS}$ as γ approaches ∞, the inverse solution $\mathbf{b}_{GLS}$ instead diverges away from the NAS solution because the regularized component $\mathbf{b}_{REG}$ blows up. Recall that the admissible range of γ value lies in the interval $(-1/r_1^2, \infty)$. For $\mathbf{b}_{GLS}$ to converge to $\mathbf{b}_{NAS}$, $\mathbf{b}_{REG}$ must go to zero, and this happens when $f_i = \sqrt{1+\gamma r_i^2}$ goes to zero, or $\gamma \to -1/r_1^2$ from the right. Note that if the processing matrix **P** used a positive exponent instead of a negative one, that is, $\mathbf{P} = \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathsf{T}}\mathbf{L}\right)^{1/2}$ instead of $\mathbf{P} = \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathsf{T}}\mathbf{L}\right)^{-1/2}$, then the diagonal element in Equation (23) would have been $f_i = 1/\sqrt{1+\gamma r_i^2}$ and $\mathbf{b}_{REG}$ would go to zero as $\gamma \to \infty$.

4.2. Generalized Tikhonov regularization inverse processing

For GTR, the inverse solution of Equation (13) (written here again for clarification)

$$\begin{bmatrix} \mathbf{X} \\ \lambda\mathbf{L} \end{bmatrix}\mathbf{b} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_n \end{bmatrix} \tag{24}$$

can be explicitly solved using one of the two methods:

(1) Transformation to the standard TR problem and transformation back [36,37].
(2) Constrain **X** and **L** to share the same loading vectors using the generalized SVD (GSVD) [36]

The solutions for these methods were first derived in the numerical analysis literature [36,37] without spectroscopic applications in mind. For a unique solution $\mathbf{b}$ to exist, $m \geq n$ (that is, $\mathbf{X}$ cannot have fewer rows than columns). This condition does not apply to the vast majority of spectroscopic data sets—unless one performs an aggressive data compression or feature selection ahead of time. However, in practice, the first method suffices—it is straightforward to implement even when $m << n$ and yields solutions that are unique *enough* (the solutions are stable, well behaved, and reproducible). The second method involves GSVD and is only of theoretical interest—it allows one to see the interrelationship between GTR and the NAS solution. First, we will describe the first method in greater detail.

4.2.1. *Transformation to standard form and back*

The method of GTR seeks to solve the linear system in Equation (24), which is equivalent to the following optimization problem:

$$\min \phi_\lambda(\mathbf{b}), \quad \text{where} \quad \phi_\lambda(\mathbf{b}) = ||\mathbf{Xb} - \mathbf{y}||^2 + \lambda^2 ||\mathbf{Lb}||^2 \tag{25}$$

Utilizing the variable transformations

$$\begin{gathered} \bar{\mathbf{X}} = \mathbf{X}\mathbf{L}_X^\dagger, \quad \bar{\mathbf{b}} = \mathbf{Lb}, \quad \bar{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}_{\text{NAS}} \\ \mathbf{L}_X^\dagger = \left[\mathbf{I}_n - \mathbf{X}_{\text{NAS}}^\dagger \mathbf{X}\right] \mathbf{L}^\dagger \end{gathered} \tag{26}$$

where $\mathbf{X}_{\text{NAS}}$ is defined in Equations (5) and (25) can be rewritten as the standard TR problem (or RR in the transformed variables):

$$\min \phi_\lambda(\bar{\mathbf{b}}), \quad \text{where} \quad \phi_\lambda(\bar{\mathbf{b}}) = ||\bar{\mathbf{X}}\bar{\mathbf{b}} - \bar{\mathbf{y}}||^2 + \lambda^2 ||\bar{\mathbf{b}}||^2 \tag{27}$$

The standard TR problem in Equation (27) is equivalent to the following augmented linear system:

$$\begin{bmatrix} \bar{\mathbf{X}} \\ \lambda \mathbf{I}_k \end{bmatrix} \bar{\mathbf{b}} = \begin{bmatrix} \bar{\mathbf{y}} \\ \mathbf{0}_k \end{bmatrix} \tag{28}$$

Recall that $\bar{\mathbf{y}}$ in Equation (28) is defined in Equation (26) as a function of $\mathbf{b}_{\text{NAS}}$, which in turn is defined in Equation (17) as $\mathbf{b}_{\text{NAS}} = \left(\mathbf{I}_n - \mathbf{QQ}^{\mathsf{T}}\right)\mathbf{g}$, where $\mathbf{g}$ is PLS model vector. As there are multiple versions of $\mathbf{b}_{\text{NAS}}$ (one version for each PLS latent vector), there are multiple versions of $\bar{\mathbf{y}}$. As a result, there are multiple solutions $\bar{\mathbf{b}}$ in Equation (28)—one for each PLS latent vector. After solving for $\bar{\mathbf{b}}$, we transform back to recover the original solution $\mathbf{b}_{\text{GTR}}$:

$$\begin{aligned} \mathbf{b}_{\text{REG}} &= \mathbf{L}_X^\dagger \bar{\mathbf{b}}, \\ \mathbf{b}_{\text{GTR}} &= \mathbf{b}_{\text{REG}} + \mathbf{b}_{\text{NAS}} \end{aligned} \tag{29}$$

The matrix $\mathbf{L}_X^\dagger$ is called the $\mathbf{X}$-*weighted generalized inverse* of $\mathbf{L}$ and can be written as

$$\mathbf{L}_X^\dagger = \left[\mathbf{I}_n - \mathbf{X}_{\text{NAS}}^\dagger \mathbf{X}\right] \mathbf{L}^\dagger = \left[\mathbf{I}_n - \left(\mathbf{X}\left(\mathbf{I}_n - \mathbf{QQ}^{\mathsf{T}}\right)\right)^\dagger \mathbf{X}\right] \mathbf{L}^\dagger \tag{30}$$

In most instances, the interferent spectra $\mathbf{L}$ have fewer samples than wavelengths ($k << n$) and Equation (30) can be used. However, if $\mathbf{L}$ has at least as many rows than columns ($k \geq n$) and has full column rank, then $\mathbf{QQ}^{\mathsf{T}} = \mathbf{I}_n$ and $\mathbf{X}_{\text{NAS}} = \mathbf{X}(\mathbf{I}_n - \mathbf{QQ}^{\mathsf{T}}) = \mathbf{0}$ in Equation (30). In this case, $\mathbf{L}_X^\dagger$ reduces to the ordinary pseudo-inverse $\mathbf{L}^\dagger$.

Although an explicit solution for $\mathbf{b}$ can be constructed using Equation (29), this equation does not shed light on how $\mathbf{b}_{\text{GTR}}$ depends upon the regularization parameter λ. This requires the GSVD of $\mathbf{X}$ and $\mathbf{L}$, which will be discussed next.

4.2.2. *Generalized singular value decomposition of* $\mathbf{X}$ *and* $\mathbf{L}$

In the GSVD discussion to follow, we require that $\mathbf{X}$ have full column rank, which implies that $m \geq n$. As this is not feasible for most spectroscopic data sets, we will augment $\mathbf{X}$ with a multiple of the identity matrix

$$\mathbf{X}_\tau = \begin{bmatrix} \mathbf{X} \\ \tau \mathbf{I}_n \end{bmatrix}$$

where τ is very small, for example, $\tau = 10^{-8}$. Although, $\tau > 0$ ensures that $\mathbf{X}_\tau$ has full column rank, it should be optimized as a tuning parameter for optimal performance. However, our interest in this augmentation is to simplify the following GSVD presentation and show the connection between GTR and NAS.

The matrix $\mathbf{X}_\tau$ now has dimension $q \times n$ where $q = m + n$. The GSVD of $\mathbf{X}_\tau$ and $\mathbf{L}$ is given by [36]:

$$\mathbf{X}_\tau = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{W}}^{-1} \quad \text{and} \quad \mathbf{L} = \mathbf{V}\hat{\mathbf{T}}\hat{\mathbf{W}}^{-1} \tag{31}$$

Both the $q \times q$ matrix $\hat{\mathbf{U}}$ and the $k \times k$ matrix $\mathbf{V}$ are orthonormal. It is useful to split $\hat{\mathbf{U}}$ into two matrices $\hat{\mathbf{U}} = [\mathbf{U}, \mathbf{U}_\perp]$, where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ and $\mathbf{U}_\perp = [\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m]$ are orthogonal complements of each other. The matrices $\hat{\mathbf{S}}$ and $\hat{\mathbf{T}}$ are "diagonal" with the following structure:

$$\hat{\mathbf{S}} = \left[\begin{array}{c|c} \mathbf{S} & \mathbf{0}_{k,q} \\ \hline \mathbf{0}_{q,k} & \mathbf{I}_q \end{array}\right], \quad \mathbf{S} = \text{diag}(s_1, \ldots, s_k), \quad \text{and}$$

$$\hat{\mathbf{T}} = \left[\mathbf{T}, \mathbf{0}_{k,q}\right], \quad \mathbf{T} = \text{diag}(t_1, \ldots, t_k)$$

where the diagonal elements s_i and t_i are non-negative and ordered in the following way: $0 \leq s_1 \leq \cdots \leq s_k \leq 1$ and $1 \geq t_1 \geq \cdots \geq t_k \geq 0$. The ratios $\theta_i = s_i/t_i$, $i = 1, \ldots, k$ are referred to as the generalized singular values and are ordered in ascending order. The rows of the $n \times n$ inverse matrix $\hat{\mathbf{W}}^{-1}$ are the loading vectors that both $\mathbf{X}$ and $\mathbf{L}$ share. The matrix $\hat{\mathbf{W}}$ can also be split into two matrices $\hat{\mathbf{W}} = [\mathbf{W}, \mathbf{W}_\perp]$, where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_k]$ and $\mathbf{W}_\perp = [\mathbf{w}_{k+1}, \ldots, \mathbf{w}_n]$ span $\mathcal{L}$ and $\mathcal{L}_\perp$, respectively. Note that the matrices $\mathbf{P}$ and $\mathbf{R}$ (associated with the SVD of $\mathbf{L}$ only) are not the same as the matrices $\mathbf{V}$ and $\mathbf{T}$. The matrices $\mathbf{P}$ and $\mathbf{V}$ span the same space but correspond to different basis sets.

4.2.3. *Generalized Tikhonov regularization inverse solution via generalized singular value decomposition*

The linear system we now work with

$$\begin{bmatrix} \mathbf{X}_\tau \\ \lambda \mathbf{L} \end{bmatrix} \mathbf{b} = \begin{bmatrix} \mathbf{y}_\tau \\ \mathbf{0}_k \end{bmatrix}, \quad \text{where} \quad \mathbf{y}_\tau = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_n \end{bmatrix}$$

is slightly different from the one in Equation (24) because of the augmentation. The solution can now be given in terms of the GSVD basis [36]:

$$\mathbf{b}_{\text{GTR}} = \mathbf{WFU}^{\mathsf{T}}\mathbf{y}_\tau + \mathbf{W}_\perp \mathbf{U}_\perp^{\mathsf{T}} \mathbf{y}_\tau = \mathbf{b}_{\text{REG}} + \mathbf{b}_{\text{NAS}} \tag{32}$$

where

$$\mathbf{F} = \text{diag}(f_1, \ldots, f_k), \quad f_i = \frac{\theta_i^2}{\theta_i^2 + \lambda^2}$$

$$\mathbf{b}_{\text{REG}} = \mathbf{WFU}^{\text{T}}\mathbf{y}_\tau = \sum_{i=1}^{k} f_i \left(\frac{\mathbf{u}_i^{\text{T}}\mathbf{y}_\tau}{s_i}\right)\mathbf{w}_i \tag{33}$$

$$\mathbf{b}_{\text{NAS}} = \mathbf{W}_\perp\mathbf{U}_\perp^{\text{T}}\mathbf{y}_\tau = \sum_{i=k+1}^{n} \left(\mathbf{u}_i^{\text{T}}\mathbf{y}_\tau\right)\mathbf{w}_i \tag{34}$$

If we were to replace $\mathbf{X}$ with $\mathbf{X}_\tau$ in Section 4.2.1, then the component-wise solutions $\mathbf{b}_{\text{REG}}$ and $\mathbf{b}_{\text{NAS}}$ in Equations (29) and (32) would be equivalent. The shrinkage of $\mathbf{b}_{\text{GTR}}$ depends entirely upon the matrix $\mathbf{F}$ in the solution component $\mathbf{b}_{\text{REG}}$. As $\lambda \to \infty$, f_i goes to zero and $\mathbf{b}_{\text{GTR}}$ approaches the NAS solution. Unlike GNAS and GLS, there is no asymptotic dichotomy associated with GTR. When $\lambda \to \infty$, both the spectral processing and inverse solutions converge to the NAS solution.

4.2.4. Generalized net analyte signal and generalized least squares spectral processing as generalized Tikhonov regularization inverse processing

In this section, we show that GNAS and GLS spectral processing can be viewed as a special case of GTR inverse processing. In Equation (25), we replace the penalty term $\lambda^2\|\mathbf{Lb}\|_2^2$ with $\|\widetilde{\mathbf{L}}\mathbf{b}\|_2^2$, where $\widetilde{\mathbf{L}} = \mathbf{P}^{-1}$ is the inverse of the $n \times n$ processing matrix associated with either GNAS or GLS. It is important to note that we are not augmenting $\mathbf{X}$ with a multiple of the interferent spectra $\mathbf{L}$. Instead, we augment $\mathbf{X}$ with the inverse of a processing matrix $\mathbf{P}$. As $\widetilde{\mathbf{L}} = \mathbf{P}^{-1}$ is a square and invertible matrix, the matrix $\widetilde{\mathbf{L}}_X^\dagger$ (the $\mathbf{X}$-weighted generalized inverse of $\widetilde{\mathbf{L}}$) simplifies to $\widetilde{\mathbf{L}}_X^\dagger = \widetilde{\mathbf{L}}^\dagger = \mathbf{P}$. As a result, the variable transformations in Equation (26) are simplified

$$\bar{\mathbf{X}} = \mathbf{XP}, \quad \bar{\mathbf{b}} = \mathbf{P}^{-1}\mathbf{b}, \quad \bar{\mathbf{y}} = \mathbf{y}$$

and the standard TR problem becomes

$$\min \phi_\lambda(\bar{\mathbf{b}}), \quad \text{where} \quad \phi_\lambda(\bar{\mathbf{b}}) = \|(\mathbf{XP})\bar{\mathbf{b}} - \mathbf{y}\|^2 + \lambda^2\|\bar{\mathbf{b}}\|^2 \tag{35}$$

Minimizing $\phi_\lambda(\bar{\mathbf{b}})$ in Equation (35) is the same as GNAS or GLS spectral processing except that we solve the modified linear system $(\mathbf{XP})\bar{\mathbf{b}} = \mathbf{y}$ by ordinary RR. From the perspective of GTR, GNAS and GLS can be viewed as *incomplete* because there is no transformation back to the original variable $\mathbf{b}$. Continuing with GTR, the transformation back to $\mathbf{b}$ is radically simplified because $\widetilde{\mathbf{L}}$ is square and invertible: $\mathbf{b}_{\text{GTR}} = \mathbf{P}\bar{\mathbf{b}}$.

4.3. Filter factor framework

We have observed that inverse processing yields a solution $\mathbf{b}_{\text{INV}}$ that can be decomposed into the sum of two terms:

$$\mathbf{b}_{\text{INV}} = \mathbf{b}_{\text{REG}} + \mathbf{b}_{\text{NAS}}$$

The role of the regularization parameters β, γ, or λ is to shrink the size of the regression vector components in $\mathbf{b}_{\text{REG}}$. The second term $\mathbf{b}_{\text{NAS}}$ is the same regardless of which inverse processing treatment (GNAS, GLS, or GTR) is invoked. The shrinkage only occurs in the $\mathbf{b}_{\text{REG}}$ term via a *filter factor* mechanism involving a matrix $\mathbf{F} = \text{diag}(f_1, \ldots, f_k)$:

$$\mathbf{b}_{\text{REG}} = \begin{cases} \mathbf{QFQ}^{\text{T}}\mathbf{g} = \sum_{i=1}^{k} f_i \left(\mathbf{q}_i^{\text{T}}\mathbf{g}\right)\mathbf{q}_i, & f_i = \begin{cases}(1-\beta)^{-1}, & \text{GNAS} \\ \sqrt{1+\gamma r_i^2}, & \text{GLS}\end{cases} \\ \mathbf{WFS}^{-1}\mathbf{U}^{\text{T}}\mathbf{y}_\tau = \sum_{i=1}^{k} f_i \left(\frac{\mathbf{u}_i^{\text{T}}\mathbf{y}_\tau}{s_i}\right)\mathbf{w}_i, \; f_i = \frac{\theta_i^2}{\theta_i^2+\lambda^2} & \text{GTR} \end{cases} \tag{36}$$

Convergence of the inverse solution to NAS solution occurs when $\mathbf{b}_{\text{REG}} = \mathbf{0}$ or $\mathbf{F} = \mathbf{0}_{k,k}$. For the different inverse processing treatments, convergence occurs when the following happens: for GNAS, when $|\beta| \to \infty$; for GLS, when $\gamma \to -1/r_1^2$; and for GTR, when $\lambda \to \infty$. Filter factor representations of $\mathbf{b}_{\text{INV}}$ by $\mathbf{b}_{\text{PLS}}$, $\mathbf{b}_{\text{PCR}}$, and $\mathbf{b}_{\text{RR}}$ have been described in [38].

5. EXPERIMENTAL AND IMPLEMENTATION

5.1. Data sets and partitions

5.1.1. Temperature data set

Twenty-two samples composed of water, ethanol, and 2-propanol were measured from 590 to 1091 nm at 1-nm intervals at temperatures 30°C, 40°C, 50°C, 60°C, and 70°C [39]. Ethanol was the analyte of interest and the spectra at temperature 30°C was used. Five out of the 22 spectra have zero analyte concentration (including two pure-component interferent spectra), and these comprised the interferent spectra $\mathbf{L}$. Another spectrum out of the 22 corresponds to a pure-component analyte spectrum and was not used. The remaining 16 spectra were split into calibration and validation sets of sample sizes 10 and 6, respectively. (The calibration and validation split of the data was the same one used in [39] but with the zero analyte samples excluded).

5.1.2. Corn data set

Eighty samples with four references—moisture, oil, protein, or starch—are measured across three different instruments—m5, mp5, and mp6 [40]. Moisture is the reference of interest. Spectra were recorded over the range of 1100–2498 nm with measurements taken at 2-nm intervals. Our task will involve calibration transfer: build a calibration model on spectra measured on the primary instrument m5 and predict samples from spectra measured on the secondary instrument mp5.

For the corn data set, there is no known default split of the data. Using the same percentage of calibration and validation samples as was used in the temperature data set: a 60%/40% calibration/validation split whereby the *first* 48 samples are set aside for the calibration set and the remaining 32 samples are used for the validation set. The interferent spectra used to form $\mathbf{L}$ are the *difference spectra* between instrument m5 and mp5 for the calibration set: the *i*th calibration spectrum from instrument m5 is subtracted from the *i*th spectrum from instrument mp5.

5.2. Computational formulas and software

Listed in Table I are the equations used to the models for spectral and inverse processing. The software used was written in MATLAB (Release 2007b) [41]. In the interest of reproducible research, the website `www.hpc.unm.edu/~andriese` will contain (in the near future) all of the MATLAB scripts and the data sets used to generate the results shown in this paper.

Table I. Equations used to form respective models for spectral and inverse processing

Method	Spectral processing	Inverse processing
NAS	Equation (5)	Equation (17)
GNAS	Equations (6) and (7)	Equations (19) and (20)
GLS	Equations (8) and (9)	Equations (22) and (23)
GTR	Equation (13)	Equations (26)–(30)

6. RESULTS AND DISCUSSION

In this section, our goal is twofold:

(1) **NAS approximation**: We want to compare the convergences of spectral and inverse processing solutions via GNAS, GLS, and GTR to solutions obtained by NAS spectral and inverse processing. When we compute the respective model vectors, the maximal rank PLS vector is used. For example, PLS is used to the solve the modified linear system $\widetilde{\mathbf{X}}\mathbf{b} = \widetilde{\mathbf{y}}$ in Equations (2) and (3) for the spectral processing cases. We select the model vector associated with maximal rank of $\widetilde{\mathbf{X}}$, that is, the rth PLS latent vector where $r = \min\{m, n\}$ with m and n being the number of rows of columns, respectively, of $\widetilde{\mathbf{X}}$. For the inverse processing cases associated with GNAS and GLS, we compute $\mathbf{b} = \mathbf{P}^{-1}\mathbf{g}$, where $\mathbf{g}$ is the least squares solution $\mathbf{X}^{\dagger}\mathbf{y}$. (Alternatively, one can use PLS and select the maximal rank model vector.) For GTR inverse processing, we compute $\mathbf{b}_{GTR} = \mathbf{L}_X^{\dagger}\bar{\mathbf{b}} + \mathbf{b}_{NAS}$, where $\bar{\mathbf{b}}$ and $\mathbf{b}_{NAS}$ are the maximal rank PLS solution vectors associated the linear system in Equation (28) and $\mathbf{X}_{NAS}\mathbf{b} = \mathbf{y}$, respectively.

(2) **Comparison of GNAS, GLS, and GTR to NAS and PLS**: We also want to compare processing solutions via GNAS, GLS, or GTR to the solutions obtained by PLS and NAS. In this case, all possible PLS latent vectors are used. The spectral and inverse processing solutions will be denoted by $\mathbf{b}_{PROC}$ and $\mathbf{b}_{INV}$, respectively. Model solutions are compared by prediction errors of the validation samples. The logarithm (base 10) of root-mean-square error of validation (RMSEV) values are tabulated for various combinations of β, γ, λ, and the number of PLS latent vectors. The number of loading vectors from the SVD of $\mathbf{L}$ is fixed to two different values for each data set as noted in the next section. We make no attempt to find the optimal regularization parameters β, γ, and λ, associated, respectively, with GNAS, GLS, and GTR, as well as the optimal number of PLS latent vectors. This attempt would be the subject of in-depth study for another paper. Again, the paper's focus is theoretical development.

With respect to building a calibration model, we opt for augmenting a vector of ones to the calibration and interferent spectra such that

$$\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{1}_m] \quad \text{and} \quad \tilde{\mathbf{L}} = [\mathbf{L}, \mathbf{1}_k]$$

The processing matrix $\mathbf{P}$ is then formed using $\tilde{\mathbf{L}}$ such that the linear system

$$(\tilde{\mathbf{X}}\mathbf{P})\tilde{\mathbf{b}} = \mathbf{y} \quad \text{or} \quad \begin{bmatrix} \tilde{\mathbf{X}} \\ \mathbf{P} \end{bmatrix} \tilde{\mathbf{b}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_k \end{bmatrix}$$

is solved for $\tilde{\mathbf{b}}$ using spectral or inverse processing. The model vector $\tilde{\mathbf{b}} = [b_1, \ldots, b_n, b_{n+1}]^T$ consists of the original vector $\mathbf{b} = [b_1, \ldots, b_n]^T$ and an offset b_{n+1}. (The augmentation causes the hyperplane-of-fit to go through the origin in an enlarged $(n+1)$-dimensional space.) The prediction $\hat{\mathbf{y}}$ of a novel spectrum $\mathbf{z} = [z_1, \ldots, z_n]^T$ involves a similar "vector-padding" by a 1 ($\tilde{\mathbf{z}} = [z_1, \ldots, z_n, z_{n+1}]^T$) and the following dot-product:

$$\hat{\mathbf{y}} = \tilde{\mathbf{z}}^T\tilde{\mathbf{b}} = \mathbf{z}^T\mathbf{b} + b_{n+1}$$

6.1. Approximation to the net analyte signal solution

In Sections 2 and 3, the spectral and inverse processing solutions $\mathbf{b}_{PROC}$ and $\mathbf{b}_{INV}$, respectively, converge to the NAS solution as their corresponding regularization parameters (β, γ, and λ) approach certain limiting values. In Table II, we summarize the limiting parameter regimes (and values used) associated with the NAS approximation.

6.1.1. *Temperature data set*

For GNAS, the processing matrix $\mathbf{P} = \mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L} = \mathbf{I}_n - \beta\mathbf{Q}\mathbf{Q}^T$ weights all of the loading vectors $\mathbf{q}_i$ the same. For effective performance, however, are all of the loading vectors necessary? In Figure 1, the singular values $r_1, \ldots, r_5$ of $\mathbf{L}$ for the temperature data set are plotted along with the variance explained for the first l ($1 \le l \le k = 5$) loading vectors. The percentage of the variance explained for the first l loading vectors is defined as $100 \times \left(\sum_{i=1}^{l} r_i^2\right) / \left(\sum_{i=1}^{k} r_i^2\right)$. The first two loading vectors capture over

Table II. Limiting parameter regimes for NAS approximation

	Spectral processing	Inverse processing
GNAS	$\beta \to 1$ ($\beta = 0.9999$)	$\beta \to \pm\infty$ ($\beta = 10,000$)
GLS	$\gamma \to \infty$ ($\gamma = 99,999,999$)	$\gamma \to c^+, c = -1/r_1^2$ ($\gamma = -0.99999999$)
GTR	$\lambda \to \infty$ ($\lambda = 10^6$)	$\lambda \to \infty$ ($\lambda = 10^6$)

Values in parentheses are the actual values used as a proxy for the limits in Figures 2 and 4.

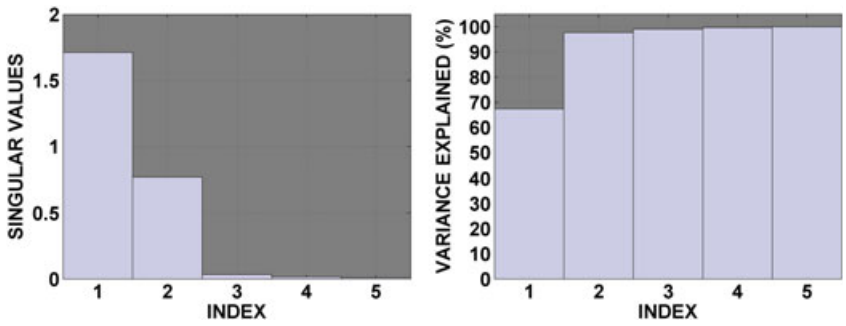

Figure 1. Temperature data singular values r_i of **L** (left) and the variance explained by the first l ($1 \leq l \leq k = 5$) loading vectors of **L** (right).

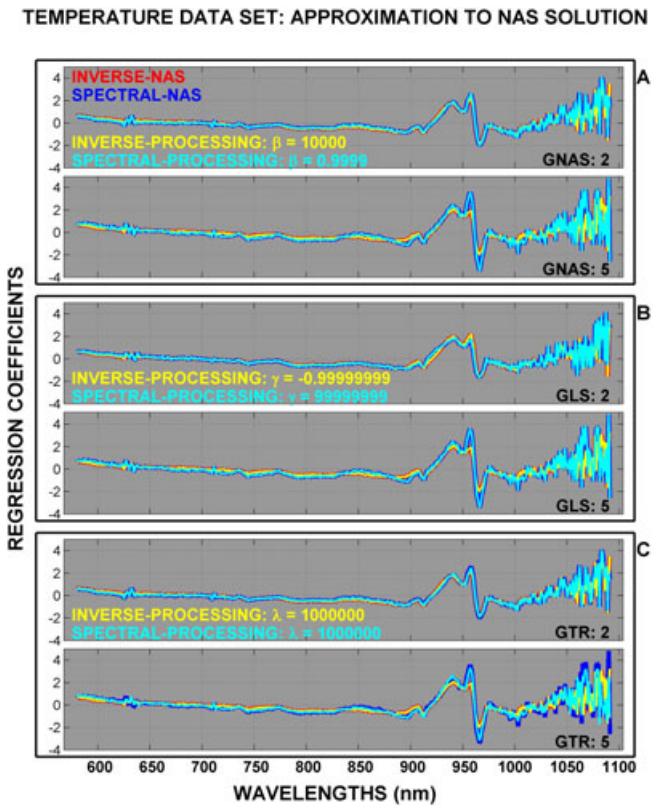

Figure 2. Temperature regression vector comparison between GNAS, GLS, GTR spectral, and inverse processing for approximating the respective NAS solutions. For A, B, and C, the top and bottom subfigures correspond to two and five loading vectors from **L** used to construct the processing matrices.

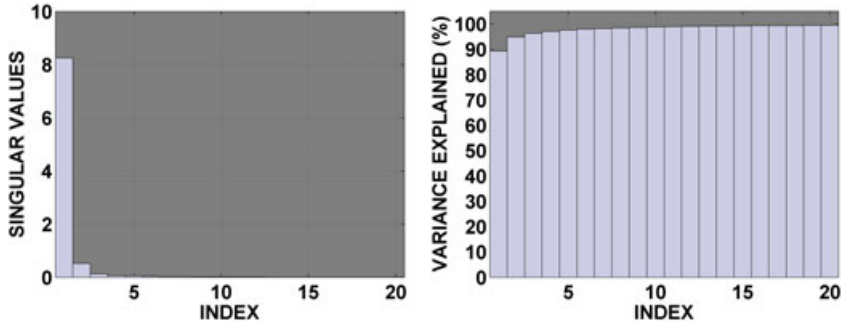

Figure 3. Corn data the singular values r_i of **L** (left) and the variance explained by the first l ($0 \leq l \leq k = 48$) loading vectors of **L** (right). For display purposes, only the results for the first 20 loading vectors are shown.

97% of the variance. Hence, we will examine the performance of NAS, GNAS, GLS, and GTR by using the SVD approximation $\mathbf{L} \approx \sum_{i=1}^{l} r_i \mathbf{h}_i \mathbf{q}_i^{\mathsf{T}}$ at $l = 2$ (rank 2 approximation of **L**) and $l = 5$ (full-rank reconstruction of **L**).

Figure 2 shows solutions (or regression vectors): GNAS (A), GLS (B), and GTR (C). Each subplot has four solutions: spectral processing by NAS (blue), inverse processing by NAS (red), respective spectral processing at the limiting value of the regularization parameter (light blue), and respective inverse processing at the limiting value (yellow). For GNAS, GLS, and GTR, the respective upper and lower subplots correspond to the solutions associated with the number of loading vectors used: $l = 2$ and $l = 5$. In the

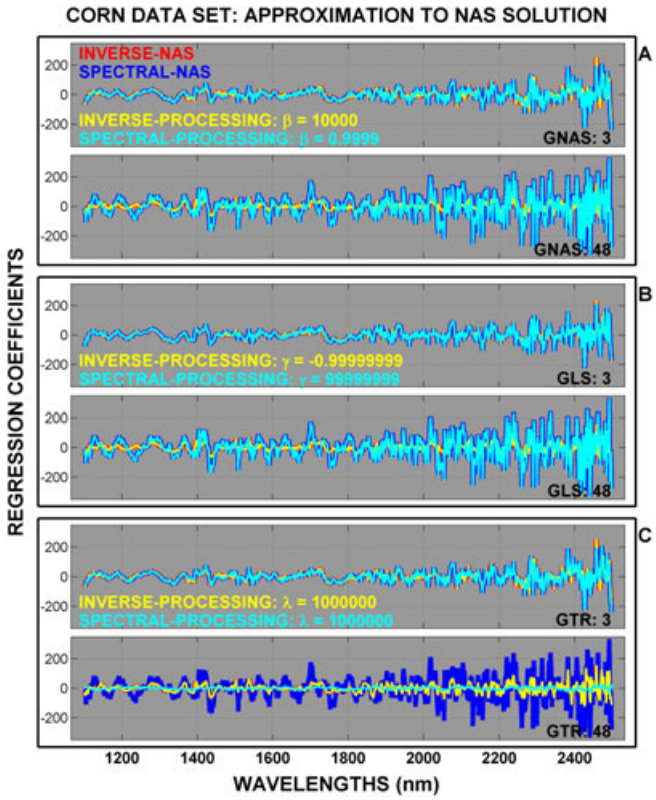

Figure 4. Corn regression vector comparison between GNAS, GLS, and GTR spectral and inverse processing for approximating the respective NAS solutions. For A, B, and C, the top and bottom subfigures correspond to 3 and 48 loading vectors from **L** used to construct the processing matrices.

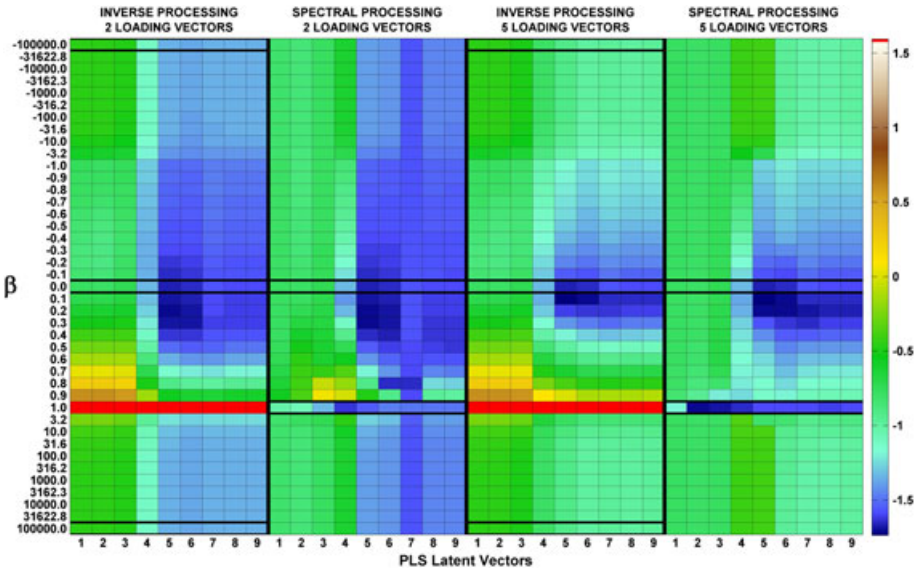

Figure 5. Temperature $\log_{10}$(RMSEV) maps for GNAS spectral and inverse processing across various regularization parameters: β on the *y*-axis and PLS latent vectors on the *x*-axis. The PLS models without processing occur at regularization parameter $\beta = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\beta = 1$ and $\beta = \pm 100,000$.

case of GLS, we have replaced **L** with its orthogonal loading vectors ($\mathbf{L} := \mathbf{Q}^T$). In this way, we have control of the limiting value of γ: $\gamma \to -1$ as opposed to $\gamma \to -1/r_1^2$, which is data set dependent.

Asymptotically, the spectral and inverse processing solutions for GNAS, GLS, and GTR all approach their respective spectral and inverse processing NAS solutions as their regularization parameters approach their limiting values—see Table II. (For inverse processing, the limiting values $\gamma = 99,999,999$ and $\gamma = -0.99999999$ were obtained by using Equation (10) with $\beta = 0.9999$ and $\beta = 10,000$, respectively.) At two loading vectors, all four regression vectors are approximately of the same size. With five loading vectors, the inverse processing regression vectors are smaller than the regression vectors associated with spectral processing.

6.1.2. *Corn data set*

In Figure 3, the singular values $r_1, \ldots, r_k$ of **L** for the corn data set are plotted along with the variance explained for the first l ($1 \leq l \leq k = 48$) loading vectors. The first three principal components capture slightly over 96% of the variance. Hence, we will examine the performance of NAS, GNAS, GLS, and GTR by using the SVD approximation $\mathbf{L} \approx \sum_{i=1}^{l} r_i \mathbf{h}_i \mathbf{q}_i^T$ at $l = 3$ and $l = 48$.

Figure 4 shows the GNAS, GLS, and GTR solutions (or regression vectors) for the corn data set as they asymptotically approach the NAS solution. Qualitatively, the behavior for the corn data set is the same as the temperature data set—the inverse and

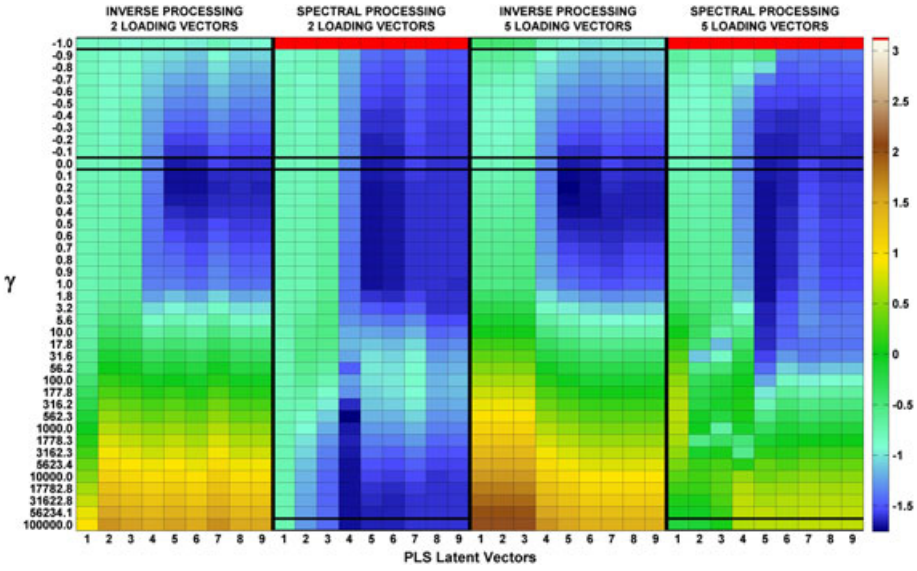

Figure 6. Temperature $\log_{10}$(RMSEV) maps for GLS spectral and inverse processing across various regularization parameters: γ on the y-axis and PLS latent vectors on the x-axis. The PLS models without processing occur at regularization parameter $\gamma = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\gamma = 100,000$ and $\gamma = -1$.

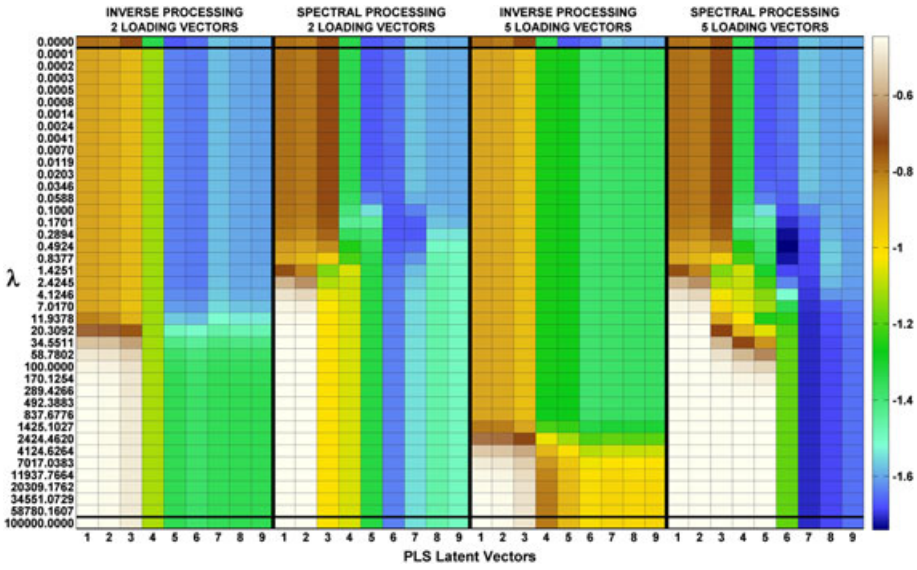

Figure 7. Temperature $\log_{10}$(RMSEV) maps for GTR spectral and inverse processing across various regularization parameters: λ on the y-axis and PLS latent vectors on the x-axis. The PLS models without processing occur at regularization parameter $\lambda = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\lambda = 100,000$.

spectral processing solutions converge to their respective NAS solutions. The exception is for GTR at full-rank—only the GTR inverse processing solution converges to the NAS inverse processing solution. The GTR spectral processing regression vector is much smaller in size than all of the other regression vectors. As with the temperature data set, the oscillatory behavior of the regression vectors of the corn data set is better behaved when using fewer loading vectors.

6.2. Temperature and corn root-mean-square error of validation behavior across β, γ, and λ

Unlike the previous section, we do not want to stress test the spectral and inverse processing solutions by pushing the regularization parameters (β, γ, and λ) to their respective limiting values, that is, removing as much interferent information as possible in order to achieve the NAS solution. Instead, this section reports on

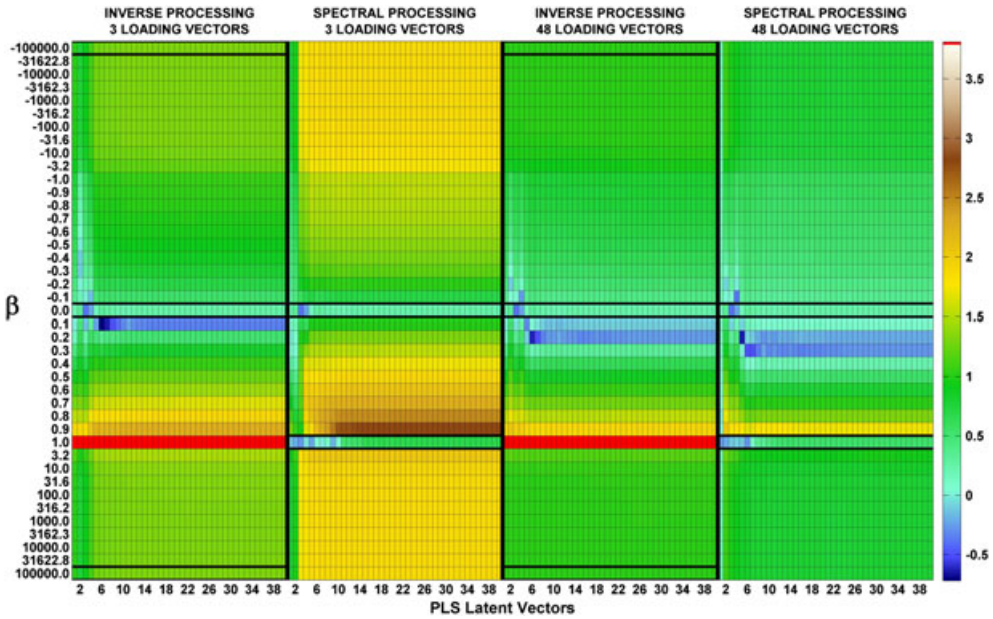

Figure 8. Corn $\log_{10}$(RMSEV) maps for GNAS spectral and inverse processing across various regularization parameters: β on the *y*-axis and PLS latent vectors on the *x*-axis. The PLS models without processing occur at regularization parameter $\beta = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\beta = 1$ and $\beta = \pm 100,000$.

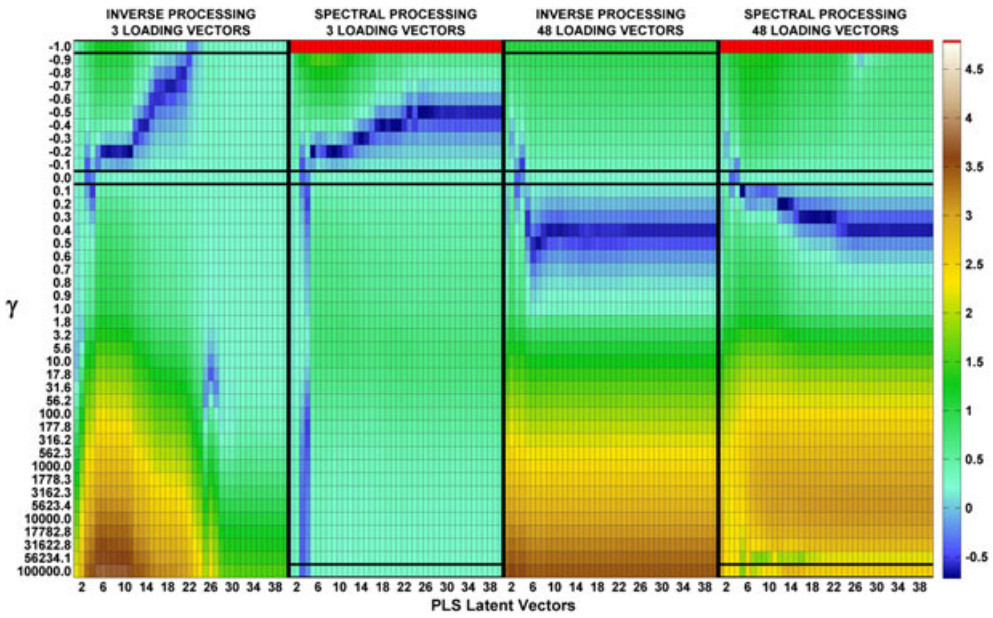

Figure 9. Corn $\log_{10}$(RMSEV) maps for GLS spectral and inverse processing across various regularization parameters: γ on the *y*-axis and PLS latent vectors on the *x*-axis. The PLS models without processing occur at regularization parameter $\gamma = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\gamma = 100,000$ and $\gamma = -1$.

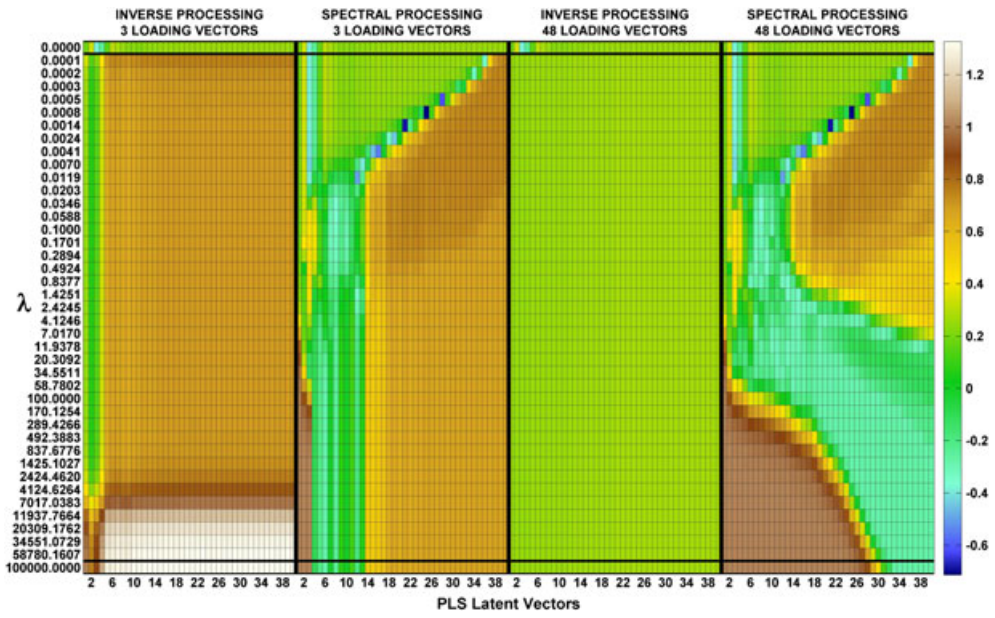

Figure 10. Corn $\log_{10}$(RMSEV) maps for GTR spectral and inverse processing across various regularization parameters: λ on the *y*-axis and PLS latent vectors on the *x*-axis. The PLS models without processing occur at regularization parameter $\lambda = 0$. Spectral and inverse GNAS processing approximate NAS processing at $\lambda = 100,000$.

studying prediction behavior when varying amounts of interferent information is retained in the processes of consisting of none (NAS), some (GNAS, GLS, and GTR), and all (PLS).

In Figures 5–10, we examine the RMSEV values on a $\log_{10}$ scale across a range of regularization parameters: $-100,000 \leq \beta \leq 100,000$, $-1 \leq \gamma \leq 100,000$, and $0 \leq \lambda \leq 100,000$. Each subplot in Figures 5–7 shows—from left to right—the RMSEV values associated with the temperature data set for inverse processing with two loading vectors, spectral processing with two loading vectors, inverse processing with five loading vectors, and spectral processing with five loading vectors. (The choice of two and five were based on the variance explained in Figure 1.) Similarly, each subplot in Figures 8–10 shows the RMSEV values associated with the corn data set for inverse processing with three loading vectors, spectral processing with three loading vectors, inverse processing with 48 loading vectors, and spectral processing with 48 loading vectors. The *x*-axis corresponds to the number of PLS latent vectors used in the construction of the inverse and spectral processing solutions. Hence, each color-coded cell corresponds to an RMSEV value for a particular number of PLS latent vectors, regularization parameter, and type of processing used (inverse or spectral) for a given number of loading vectors.

For GNAS inverse processing and GLS spectral processing, the red cells are for $\beta = 1$ and $\gamma = -1$, respectively. For GNAS inverse processing, $\mathbf{b}_{REG}$ is undefined for $\beta = 1$ in Equation (19), whereas for GLS spectral processing, the filter factor f_i is undefined for $\gamma = -1$ in Equation (9). Each subplot also shows two horizontal bands across regularization parameters corresponding to the PLS and NAS solutions: $\beta = \gamma = \lambda = 0$ for ordinary PLS solutions (no interferent used), $\beta = 1$ for GNAS spectral processing approximation to NAS, $\beta = \pm 100,000$ for GNAS inverse processing approximation to NAS, $\gamma = 100,000$ for GLS spectral processing approximation to NAS, $\gamma = -1$ for GLS inverse processing approximation to NAS, and $\lambda = 100,000$ for spectral and inverse processing approximation to NAS.

For the temperature data set, we highlight some of the general trends. Relative to the RMSEV values obtained by ordinary PLS, improved performance appears possible by removing some interferent information. Similarly, compared with NAS processing, improved performance is possible including some interferent information. For GNAS and GLS (the schemes that do not involve augmentation), using fewer loading vectors of **L** results in an increase in the number of cells that have lower RMSEV—a likely consequence of the benefit derived by shedding the spurious noise associated with the latter set of loading vectors. On the other hand, GTR derives a modest boost in performance when using more loading vectors with spectral processing. For the corn data set, the trends are similar.

7. CONCLUSION

For spectral and inverse processing using GNAS and GLS, the respective β and γ tuning parameters have asymptotic dichotomy for convergence to the orthogonal NAS. Conversely, as the tuning parameter $\lambda \to \infty$ for either GTR spectral or inverse processing, GTR converges to orthogonal NAS. From the results presented, GTR is best used in the spectral processing format. It should be noted that while PLS is used as the second regularization for GTR spectral processing, that is, the latent vectors serve as the second tuning parameter, using a more continuous regularization parameter can provide different results [24,25].

The interferent matrix **L** can include any artifact that the model needs to be desensitized to as in the orthogonal NAS processing examples previously noted [10–21]. It may be that improvements are possible in these references by using oblique adjustment via GNAS, GLS, or GTR. Regardless of the approach, it is generally found that when interferent spectra are isolated from the calibration samples with reference values and weighted by β, γ, or λ, fewer calibration reference samples are needed to form the

calibration model [9,22] as well as the possibility of complete elimination of reference samples [23].

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant no. CHE-1111053 (co-funded by MPS Chemistry and the OCI Venture Fund) and is gratefully acknowledged by the authors.

REFERENCES

1. Næs T, Isaksson T, Fern T, Davies T. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications: Chichester, UK, 2002.
2. Hastie TJ, Tibshirani RJ, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd edn). Springer-Verlag: New York, 2009.
3. Kalivas JH. Calibration methodologies. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis Vol. 3,* Brown S, Tauler R, Walczak B (eds-in-chief). Elsevier: Amsterdam, 2009; 1–32.
4. Brown CD. Discordance between net analyte signal theory and practical multivariate calibration. *Anal. Chem.* 2004; **76**: 4364–4373.
5. DiFoggio R. Desensitizing models using covariance matrix transforms or counter-balanced distortion. *J. Chemom.* 2005; **19**: 203–215.
6. Zeaiter M, Roger J-M, Bellon-Maurel V. Robustness of models developed by multivariate calibration. Part II: the influence of preprocessing methods. *Trends Anal. Chem.* 2005; **24**: 437–445.
7. Zeaiter M, Rutledge D. Preprocessing methods. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis Vol. 3,* Brown S, Tauler R, Walczak B (eds-in-chief). Elsevier: Amsterdam, 2009; 121–231.
8. Martens H, Høy M, Wise BM, Bro R, Brockhoff PB. Pre-whitening of data by covariance-weighted pre-processing. *J. Chemom.* 2003; **17**: 153–165.
9. Shi Z, Cogdill RP, Martens H, Anderson CA. Optical coefficient-based multivariate calibration on near-infrared spectroscopy. *J. Chemometrics* 2009; **24**: 288–299.
10. Andrew A, Fearn T. Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation. *Chemom. Intell. Lab. Syst.* 2004; **72**: 51–56.
11. Westerhaus MO. Improving repeatability of NIR calibrations across instruments. In *Proceedings of the Third International Near Infrared Spectroscopy Conference*, Biston R, Bartiaux-Thill N (eds). Agriculture Research Centre Publishing: Gembloux, Belgium, 1991; 671–674.
12. Zhu Y, Fearn T, Samuel D, Dhar A, Hameed O, Bown SG, Lovat LB. Error removal by orthogonal subtraction (EROS): a customised pre-treatment for spectroscopic data. *J. Chemom.* 2008; **22**: 130–134.
13. Sun J. Statistical analysis of NIR data: data pretreatment. *J. Chemom.* 1997; **11**: 525–532.
14. Vogt F, Mizaikoff B. Fault-tolerant spectroscopic data evaluation based on extended principal component regression correcting for spectral drifts and uncalibrated spectral features. *J. Chemom.* 2003; **17**: 660–665.
15. Ferré J, Brown SR. Reduction of model complexity by orthogonalization with respect to non-relevant spectral changes. *Appl. Spectrosc.* 2001; **55**: 708–714.
16. Roger JM, Chauchard F, Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* 2003; **66**: 191–204.
17. Shi Z, Anderson CA. Scattering orthogonalization of near-infrared spectra for analysis of pharmaceutical tablets. *Anal. Chem.* 2009; **81**: 1389–1396.
18. Chen Z, Morris J, Martin E. Extracting chemical information from spectral data with multiplicative light scattering effects by optical path-length estimation and correction. *Anal. Chem.* 2006; **78**: 7674–7681.
19. Soyemi OO, Landry MR, Yang Y, Idwasi P, Soller BR. Skin color correction for tissue spectroscopy: demonstration of a novel approach with tissue-mimicking phantoms. *Appl. Spectrosc.* 2005; **59**: 237–244.
20. Yang Y, Shoer L, Soyemi OO, Landry MR, Soller BR. Removal of analyte-irrelevant variations in near-infrared tissue spectra. *Appl. Spectrosc.* 2006; **60**: 1070–1077.
21. Yang Y, Soyemi OO, Landry MR, Soller BR. Noninvasive in vivo measurement of venous blood pH during exercise using near-infrared reflectance spectroscopy. *Appl. Spectrosc.* 2007; **61**: 223–229.
22. Hansen PW. Pre-processing method minimizing the need for reference analysis. *J. Chemom.* 2001; **15**: 123–131.
23. Ottaway J, Farrell JA, Kalivas JH. Spectral multivariate calibration without laboratory prepared or determined reference analyte values. *Anal. Chem.* 2013; **85**: 1509–1516.
24. Kalivas JH, Siano GS, Andries E, Goicoechea HC. Calibration maintenance and transfer using Tikhonov regularization approaches. *Appl. Spectrosc.* 2009; **63**: 800–809.
25. Kalivas JH. Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance. *J. Chemom.* 2012; **26**: 218–230.
26. Marbach R. A new method for multivariate calibration. *J. Near Infrared Spectrosc.* 2005; **13**: 241–254.
27. Boulet J-C, Roger JM. Improvement of direct calibration in spectroscopy. *Anal. Chim. Acta* 2010; **668**: 130–136.
28. Lawson CL, Hanson RJ. *Solving Least Squares Problems*. Prentice Hall Press: Englewood Cliffs, NJ, 1974.
29. 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.
30. Kunz MR, Ottaway J, Kalivas JH, Andries E. Impact of standardization sample design on Tikhonov regularization variants for spectroscopic calibration maintenance and transfer. *J. Chemom.* 2010; **24**: 218–229.
31. Haaland DM, Melgaard DK. New augmented classical least squares for improved quantitative spectral analyses. *Vib. Spectrosc.* 2002; **29**: 171–175.
32. Haaland DM, Melgaard DK. New prediction augmented classical least squares (PACLS): application to unmodeled interferents. *Appl. Spectrosc.* 2000; **54**(9): 1303–1312.
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. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM Press: Philadelphia, PA, 1998.
37. Elden L. A weighted pseudoinverse, generalized singular values and constrained least squares problems. *BIT* 1982; **22**: 487–501.
38. Andries E, Kalivas JH. Multivariate calibration leverages and spectral *F*-ratios via the filter factor representation. *J. Chemom.* 2010; **24**: 249–260.
39. Wülfert F, Kok WT, Smilde AK. Near-infrared data with temperature effects. *Anal. Chem.* 1998; **70**: 1761–1767.
40. NIR of corn samples for standardization benchmarking. Eigenvector Research Incorporated. http://www.eigenvector.com/data/Corn/index.html.
41. The MathWorks (www.mathworks.com), Natick, MA 01760-2098. USA.

APPENDIX A: MATHEMATICAL EXPRESSION OF GLS PROCESSING MATRIX

The strategy for spectral processing in [8,9] involves the post-multiplication **X** by a covariance matrix involving interferent spectra. Borrowing from the notation of [8], suppose **X** can be rewritten as a mixture of pure and interferent spectra

$$\mathbf{X} = \mathbf{C}_S\mathbf{S} + (\mathbf{C}_L\mathbf{L} + \mathbf{E})$$

where $\mathbf{C}_S$ and $\mathbf{C}_L$ represent the analyte concentrations for the pure-component spectra $\mathbf{S}$ and interferent spectra $\mathbf{L}$, respectively, and $\mathbf{E}$ represents the unknown residuals in $\mathbf{X}$. The matrix $\mathbf{C}_L\mathbf{L} + \mathbf{E}$ is the contribution to the spectral mixture due to unwanted chemical and physical interferents and noise. As a result, the expression

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{C}_L\mathbf{L} + \mathbf{E}) = \mathbf{L}^{\mathrm{T}}\text{cov}(\mathbf{C}_L)\mathbf{L} + \text{cov}(\mathbf{E}) \tag{A1}$$

designates the "uncertainty" covariance. As in practice one does not know cov($\mathbf{C}_L$), Martens *et al*. [8] replaced it with a simplified covariance model $\gamma\mathbf{L}^{\mathrm{T}}\mathbf{L}$. Likewise, cov($\mathbf{E}$) is unknown in practice and replaced with $\mathbf{I}_n$, that is, a covariance matrix where all wavelengths are uncorrelated and have unit variance. As a result, Equation (A1) simplifies to

$$\boldsymbol{\Sigma} = \mathbf{L}^{\mathrm{T}}\text{cov}(\mathbf{C}_L)\mathbf{L} + \text{cov}(\mathbf{E}) = \mathbf{I}_n + \gamma\mathbf{L}^{\mathrm{T}}\mathbf{L}$$

The processing matrix now becomes

$$\mathbf{P} = \boldsymbol{\Sigma}^{-1/2} = \left(\mathbf{I}_n + \gamma\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)^{-1/2} \tag{A2}$$

where the exponent –1/2 is a consequence of the weighted least squares "pre-whitening" procedure whereby the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ is split into equal matrices—see [8, Appendix II] for details.

APPENDIX B: OUTER PRODUCT OF BLOCK MATRICES

Let the $n \times n$ matrix $\mathbf{C}$ be split into sub-matrices such that $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$, $\mathbf{C}_1 = [\mathbf{c}_1, \ldots, \mathbf{c}_k]$ and $\mathbf{C}_2 = [\mathbf{c}_{k+1}, \ldots, \mathbf{c}_n]$. Furthermore, let

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0}_{k,q} \\ \mathbf{0}_{q,k} & \mathbf{D}_2 \end{bmatrix}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are matrices of dimension $k \times k$ and $q \times q$ where $q = n - k$. Then the matrix $\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^{\mathrm{T}}$ can be expressed in terms of an outer product:

$$\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}^{\mathrm{T}} = [\mathbf{C}_1, \mathbf{C}_2]\begin{bmatrix} \mathbf{D}_1 & \mathbf{0}_{k,q} \\ \mathbf{0}_{q,k} & \mathbf{D}_2 \end{bmatrix}[\mathbf{C}_1, \mathbf{C}_2]^{\mathrm{T}} = \mathbf{C}_1\mathbf{D}_1\mathbf{C}_1^{\mathrm{T}} + \mathbf{C}_2\mathbf{D}_2\mathbf{C}_2^{\mathrm{T}}$$

If $\mathbf{C} = \hat{\mathbf{Q}} = [\mathbf{Q}, \mathbf{Q}_\perp]$ (using the SVD notation from Section 2.2) is orthonormal, then the inverse (or pseudo-inverse) of $\mathbf{A}$ can also be written as an outer product sum:

$$\mathbf{A}^{\dagger} = \left(\hat{\mathbf{Q}}\mathbf{D}\hat{\mathbf{Q}}^{\mathrm{T}}\right)^{\dagger} = \hat{\mathbf{Q}}\mathbf{D}^{\dagger}\hat{\mathbf{Q}}^{\mathrm{T}} = [\mathbf{Q}, \mathbf{Q}_\perp]\begin{bmatrix} \mathbf{D}_1 & \mathbf{0}_{k,q} \\ \mathbf{0}_{q,k} & \mathbf{D}_2 \end{bmatrix}^{\dagger}$$
$$[\mathbf{Q}, \mathbf{Q}_\perp]^{\mathrm{T}} = \mathbf{Q}\mathbf{D}_1^{\dagger}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{D}_2^{\dagger}\mathbf{Q}_\perp^{\mathrm{T}}$$

Moreover, if the diagonal elements of $\mathbf{D}_1$ and $\mathbf{D}_2$ consist of positive entries, then any matrix power r (positive or negative, fractional or integer) of $\mathbf{A}$ can be expressed as

$$\mathbf{A}^r = \left(\hat{\mathbf{Q}}\mathbf{D}\hat{\mathbf{Q}}^{\mathrm{T}}\right)^r = \hat{\mathbf{Q}}\mathbf{D}^r\hat{\mathbf{Q}}^{\mathrm{T}} = [\mathbf{Q}, \mathbf{Q}_\perp]\begin{bmatrix} \mathbf{D}_1 & \mathbf{0}_{k,q} \\ \mathbf{0}_{q,k} & \mathbf{D}_2 \end{bmatrix}^r$$
$$[\mathbf{Q}, \mathbf{Q}_\perp]^{\mathrm{T}} = \mathbf{Q}\mathbf{D}_1^r\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{D}_2^r\mathbf{Q}_\perp^{\mathrm{T}}$$

APPENDIX C: GENERALIZED NET ANALYTE SIGNAL AND GENERALIZED LEAST SQUARES PROCESSING MATRICES

The decomposition of the GNAS or GLS processing matrix requires two facts from linear algebra involving the orthonormal matrix $\hat{\mathbf{Q}} = [\mathbf{Q}, \mathbf{Q}_\perp]$: $\mathbf{I}_n = \hat{\mathbf{Q}}\hat{\mathbf{Q}}^{\mathrm{T}} = \mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}$ and

$$\left(\hat{\mathbf{Q}}\begin{bmatrix} \mathbf{D}_1 & \mathbf{0}_{k,q} \\ \mathbf{0}_{q,k} & \mathbf{D}_2 \end{bmatrix}\hat{\mathbf{Q}}^{\mathrm{T}}\right)^r = \mathbf{Q}\mathbf{D}_1^r\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{D}_2^r\mathbf{Q}_\perp^{\mathrm{T}}$$

from Appendix B where the diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ have positive entries.

The GNAS processing matrix $\mathbf{P}$ in Equation (6) and its inverse can now be written as

$$\begin{aligned}\mathbf{P} &= \left(\mathbf{I}_n - \beta\mathbf{L}^{\dagger}\mathbf{L}\right) \\ &= \mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}} - \beta\mathbf{Q}\mathbf{Q}^{\mathrm{T}} \\ &= \mathbf{Q}(\mathbf{I}_k - \beta\mathbf{I}_k)\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}} \\ &= (1-\beta)\mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}\end{aligned}$$

and

$$\mathbf{P}^{-1} = \left((1-\beta)\mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}\right)^{-1} = \frac{1}{(1-\beta)}\mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}$$

Similarly, the GLS processing matrix $\mathbf{P}$ in Equation (8) and its inverse can be written as

$$\begin{aligned}\mathbf{P} &= (\mathbf{I}_n + \gamma\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1/2} \\ &= \left(\mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}} + \gamma\mathbf{Q}\mathbf{R}^2\mathbf{Q}^{\mathrm{T}}\right)^{-1/2} \\ &= \left(\mathbf{Q}\left(\mathbf{I}_k + \gamma\mathbf{R}^2\right)\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}\right)^{-1/2} \\ &= \mathbf{Q}(\mathbf{I}_k + \gamma\mathbf{R}^2)^{-1/2}\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}} \\ &= \mathbf{Q}(\mathbf{I}_k + \gamma\mathbf{R}^2)^{-1/2}\mathbf{Q}^{\mathrm{T}} + \left(\mathbf{I}_n - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\right) \\ &= \mathbf{I}_n - \mathbf{Q}\left[\mathbf{I}_k - (\mathbf{I}_k + \gamma\mathbf{R}^2)^{-1/2}\right]\mathbf{Q}^{\mathrm{T}}\end{aligned}$$

and

$$\mathbf{P}^{-1} = \left(\mathbf{Q}(\mathbf{I}_k + \gamma\mathbf{R}^2)\mathbf{Q}^{\mathrm{T}} + \mathbf{Q}_\perp\mathbf{Q}_\perp^{\mathrm{T}}\right)^{1/2} = \mathbf{I}_n - \mathbf{Q}\left[\mathbf{I}_k - (\mathbf{I}_k + \gamma\mathbf{R}^2)^{1/2}\right]\mathbf{Q}^{\mathrm{T}}$$

