Journal of Biomedical Optics 7(1), 130–147 (January 2002)

On Wiener filtering and the physics behind statistical modeling

Ralf Marbach
VTT Electronics
Kaitovayla 1
90571 Oulu, Finland

Abstract. The closed-form solution of the so-called statistical multivariate calibration model is given in terms of the pure component spectral signal, the spectral noise, and the signal and noise of the reference method. The "statistical" calibration model is shown to be as much grounded on the physics of the pure component spectra as any of the "physical" models. There are no fundamental differences between the two approaches since both are merely different attempts to realize the same basic idea, viz., the spectrometric Wiener filter. The concept of the application-specific signal-to-noise ratio (SNR) is introduced, which is a combination of the two SNRs from the reference and the spectral data. Both are defined and the central importance of the latter for the assessment and development of spectroscopic instruments and methods is explained. Other statistics like the correlation coefficient, prediction error, slope deficiency, etc., are functions of the SNR. Spurious correlations and other practically important issues are discussed in quantitative terms. Most important, it is shown how to use *a priori* information about the pure component spectra and the spectral noise in an optimal way, thereby making the distinction between statistical and physical calibrations obsolete and combining the best of both worlds. Companies and research groups can use this article to realize significant savings in cost and time for development efforts. [DOI: 10.1117/1.1427051]

Keywords: multivariate; calibration; chemometrics; pure component spectrum; Wiener filter; signal-to-noise ratio.

Paper JBO-001017 received Feb. 26, 2001; revised manuscript received July 3, 2001; accepted for publication July 6, 2001.

1 Introduction

Biomedical and other optical measurements are often based on so-called multivariate calibration. For this, an instrument measures a set of multiple input signals first, e.g., light absorbance values at different optical wavelengths, and then an algorithm is used to transform the many input numbers into one user-desired output number. Multivariate calibration, also known as (aka) chemometrics, is the process of determining that algorithm. The most popular calibration method is linear regression of the so-called ''statistical'' or ''inverse'' model. This approach, however, so far has suffered from lack of understanding of the underlying physics and thus has been considered a statistical or ''soft-modeling'' tool.

In this article, the closed-form solution of the statistical calibration model is given as a function of the pure component spectral signal, the spectral noise, and the signal and noise of the reference method. The solution is a fairly complex formula which does, however, provide a wealth of practical benefits in several ways. First, it can be used to speed up the convergence against the desired, optimum Wiener filter. In particular, the effects of spurious correlations and reference noise can be eliminated. Second, it can be used to guarantee specificity. Third, it makes the calibration process fully transparent. Also, all relevant measures of prediction quality can be shown to be functions of a single basic quantity, viz., the application-specific signal-to-noise ratio (SNR).

It turns out that current chemometrics' practices can be improved in many ways. Practical pieces of ''how-to'' information that can be gathered from this article include how to make effective use of *a priori* knowledge about the pure component spectrum and/or the spectral noise; how to interpret a prediction slope smaller than one and how to ''correct'' it; how to effectively deal with spurious correlations; how to make conscious decisions about whether or not to utilize unspecific correlations; how to build up closed-loop communications between the hardware people and the application developers in a company; how to select a ''good'' wavelength range; how to define the coordinate system that breaks the one multivariate measurement down into many univariate ones; how to effectively rank noise sources; how to measure the quality of a measurement system and quantify progress made and progress needed; and very importantly, how to reduce the number of expensive calibration experiments.

The chemometrics field encompasses a wide variety of applications, each with a different set of practical problems. Without any ranking or claim for completeness, this author's list of encountered calibration problems includes (1) instation-

Address all correspondence to Ralf Marbach. Tel: +358 8 551 2249; Fax: +358 8 551 2320; E-mail: ralf.marbach@vtt.fi

1083-3668/2002/$15.00 © 2002 SPIE

arity of spectral response; (2) nonlinearity of spectral response; (3) outliers; (4) ill-posed spectral response; (5) low spectral SNR_x; (6) low reference SNR_y; (7) unknown shape of the pure component response spectrum; (8) spurious correlations and/or overfitting; (9) unspecific correlations; and (10) bad quality of the estimate of the future spectral noise. In this article we address problems (4)–(10), and touch upon (3), but do not address (1) and (2) at all. The latter means, mathematically speaking, that it is assumed throughout the article that the linear and stationary model, Eq. (1), is valid. Practically speaking, it means that the results reported will have a direct and major impact on many ill-posed measurement applications, where the signals are too small to cause any nonlinearity and the samples are stationary, e.g., many biomedical infrared (IR) applications; whereas in other applications, notably those in industrial process control, potential nonlinearity/instationarity problems first have to be solved before they can reap the full benefit.

[We briefly define nonlinearity and instationarity of spectral response here by citing a paper by Schmitt and Kumar.[1] The authors give quantitative expressions for the effective optical pathlength in diffuse reflection experiments using fiber probes. For example, in the case of large fiber separation, $L_{\text{eff}} \cong \sqrt{3\mu_{st}/4\mu_a}\,\rho$ where μ_{st} and μ_a are the transport scattering and absorbance coefficient (mm^{-1}) of the sample and ρ is the fiber separation (mm). Thus, the measured ''absorbance,'' $-\log R_{\text{diff}} \propto \sqrt{\mu_a \mu_{st}}$ is *nonlinear* in μ_a and can be *nonstationary over time* with changes in the scatter coefficient μ_{st}. A hardware design method by which to minimize these effects is also given by those authors.[2] Other practical methods to minimize nonlinearity and nonstationarity are physical reduction of the sample variability to the extent that the linear and stationary approximation, Eq. (1), becomes valid and various mathematical data pretreatments, see, e.g., Ref. 3.]

Publications in the fields of statistics, chemometrics, and time-signal processing were reviewed for this article. The closed-form solution, Eq. (12), seems to be new. Many publications on time-signal processing were found to be not directly relevant to chemometrics because the electronic noise is usually assumed to be ''white'' (not correlated from one sampling moment in time to the next), i.e., its covariance matrix is a (scaled) identity matrix. In chemometrics, however, the covariance matrix of the spectral noise (defined below) is typically highly structured due to the correlations between wavelengths. Likewise, many of the publications in the statistics field are also not directly relevant, because chemometrics is a physical measurement problem, not a problem of finding statistical relationships. The following literature review will focus on publications with emphasis in two areas: (a) the use of knowledge about the pure component spectrum in the context of ''statistical'' calibration and (b) the effect of noise in the spectra, i.e., on the *right* side variables of the regression model, Eq. (1). We start with the latter.

There seems to be no standard method used by statisticians to deal with noise in the right-side variables, except for the univariate case.[4] The fact that the estimates of the slope coefficients are decreased by right-side noise in both the uni- and multivariate cases has been known for many years in statistics,[5] however, the effect seems to be of little importance to most statistical applications. On the other hand, the signal-processing community has recently developed interest in the thematic and has started to use the total-least-squares technique as a vehicle to incorporate right-side noise,[6] and at least one contribution by the chemometrics field has been made.[7] The chemometrical literature contains a series of papers on the net analyte signal (NAS), which is defined as that part of the pure component spectrum that is orthogonal to all other spectra.[8,9] The NAS concept points in the right direction, i.e., it tries to quantify how much of the pure component spectrum is useful in calibration, but it suffers from a basic insufficiency. In a Gedanken experiment, the more spectra are included in the list of ''other spectra,'' the smaller the NAS will get, even if the spectra included have very small amplitudes. The severeness of spectral overlap is obviously governed by both the spectral shape and the magnitude of the interfering component, so NAS is incomplete. It will be shown below that NAS is basically identical to the ''classical'' model, and the inferiority to the Wiener filter will be interpreted in terms of the assumptions that these approaches implicitly make about the spectral SNR_x. Still, NAS and related concepts have been successfully applied in a number of different applications, including chemometrical calibrations,[10,11] estimation of sinusoidal frequencies in signal processing[12] and hyperspectral image processing.[13]

A summary of various empirically proposed measures of SNR in chemometrics has recently been given,[14] however, these definitions focus exclusively on instrument noise in the spectra. It will be shown below that, in order to arrive at the closed-form solution, the definition of ''spectral noise'' must include both instrument noise and the interfering spectra from the other components, and treat them as indistinguishable. Also, the definition of reference ''signal'' and ''noise'' must be made in the particular way given in Eq. (9) below.

In order to simplify the discussion and to assign physical units to the quantities involved, the biomedical application of infrared blood glucose sensing will be used as an example, with the glucose concentration measured in units of (mg/dL) and the infrared spectra measured in units of absorbance (AU). The discussion, however, is not restricted in any way to glucose sensing or to IR spectroscopy but applies to all multichannel measurement systems in which noisy input data are measured and linearly calibrated to produce an output number.

2 Notation

Upper case bold letters denote matrices (e.g., $\mathbf{X}$) and lower case bold letters denote column vectors (e.g., $\mathbf{b}$). The index in $\mathbf{X}_{(m\times k)}$ means that the matrix has m rows and k columns. The following indices will be used: m is the number of calibration spectra, k is the number of channels or ''pixels'' per spectrum, and n is the number of future prediction spectra. $\mathbf{X}^T$ denotes the transpose; $(\mathbf{X}^T\mathbf{X})^{-1}$ an inverse; $\mathbf{X}^+$ the Moore–Penrose inverse; $\mathbf{I}$ the identity matrix; $\mathbf{1}$ a vector of ones, $(1,1,1,...,)^T$; $\mathbf{0}$ a vector of zeros, $(0,0,0,...,)^T$; $\|\mathbf{b}\|$ the Euclidean length of vector $\mathbf{b}$; and $a \equiv b$ means a is equal to b by definition.

Useful terminology for describing calibration and prediction errors is introduced schematically in Figure 1 where a straight line has been least squares fitted through the prediction scatter plot *a posteriori*. Scatter plots, by convention, show the concentrations measured by the infrared method on the y axis and the ''true'' concentrations measured by a refer-

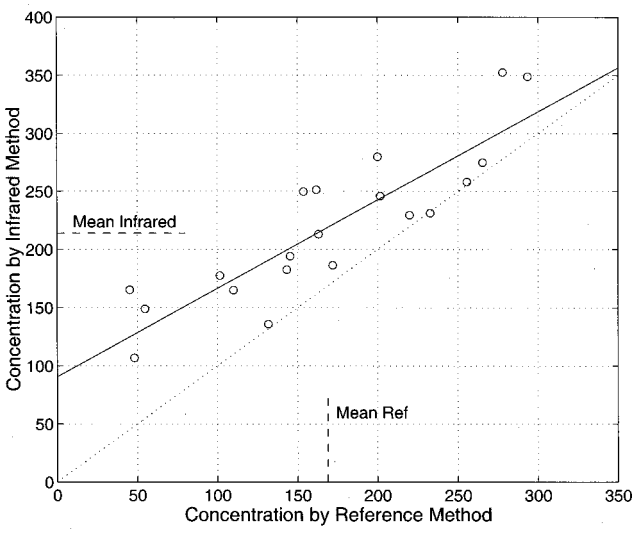

Fig. 1 Schematic of a scatter plot with an identity line (dotted) and an *a posteriori* least-squares fitted line (solid). In this example, the bias is 45 (arbitrary concentration units) and the slope of the fitted line is 0.7, with the line rotated around the point where the two means meet.

ence method on the *x* axis. The terms are the following.

1. The *bias error* (mg/dL) is the difference between the average predicted concentration and the average reference concentration; the bias, by mathematical definition, is zero for the calibration fit and the goal is to keep it zero during future predictions.
2. The *slope* (unitless) is the slope of the *a posteriori* least-squares fitted line and is almost always smaller than 1, a fact which is referred to as *slope deficiency*.
3. The *slope error* (mg/dL) of a particular prediction sample is the difference between the identity line and the bias-corrected *a posteriori* fitted line at that sample's reference concentration value; the slope error causes the above-average concentrations to be consistently underestimated and vice versa; the slope error of the whole prediction data set is the root sum of squares (RSS) over all samples.
4. The *scatter error* (mg/dL) of a particular prediction sample is the difference between the predicted value and the *a posteriori* line; the scatter error of the whole prediction data set is the RSS over all samples.

Mathematical definitions for the terms will be given below. Suffice it to say here that bias, slope, and scatter errors can all be easily measured individually by fitting the *a posteriori* line through a given scatter cloud; that the total prediction error, aka $\text{PRESS}^{1/2}$, is the root sum of squares of the bias, slope, and scatter errors; and that the slope error decreases and the scatter error increases with an increase of the slope.

3 Theory

Let us assume a set of blood samples is available for calibration. In the calibration experiment, *m* infrared spectra with *k* channels each are measured. Simultaneously, using an established clinical analysis reference method, the glucose concentrations of the blood samples are also determined. The following linear regression equation is the so-called multivariate ''statistical'' calibration model:

$$\mathbf{y}_R = \mathbf{X}\cdot\mathbf{b} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}_{\mathbf{R}(mx1)}$ is the vector of glucose concentration references (in units of mg/dL), $\mathbf{X}_{(mxk)}$ is the matrix of infrared calibration spectra (AU), $\mathbf{b}_{(kx1)}$ is the regression vector (mg/dL/AU), and $\mathbf{e}_{(mx1)}$ is the error vector (mg/dL). (The term ''multivariate'' is commonly used in the chemometric community. Readers from different backgrounds should be aware that the majority of spectrometric applications are actually better described by what they call ''multidimensional'' or ''multichannel'' measurements, because the input data comes from a physical measurement and not from a statistical selection of variables.)

The task is to find a solution for the regression vector or *b* vector **b** which minimizes the length of the error vector **e** and performs well in future predictions.

The standard procedure is to, first, mean center the calibration data,

$$\tilde{\mathbf{X}} \equiv \mathbf{X} - \mathbf{1}_{(mx1)}\cdot\overline{\mathbf{x}}^T, \tag{2}$$

$$\tilde{\mathbf{y}}_R \equiv \mathbf{y}_R - \overline{y_R}, \tag{3}$$

where $\overline{\mathbf{x}}^T$ and $\overline{y_R}$ denote the mean infrared spectrum and the mean glucose reference concentration of the calibration data set, respectively; and then, second, to estimate **b** from the least-squares (LS) solution,

$$\hat{\mathbf{b}} = \tilde{\mathbf{X}}^{+}\tilde{\mathbf{y}}_R = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{y}}_R\,. \tag{4}$$

[Note that Eq. (4) assumes $\tilde{\mathbf{X}}$ to have full column rank, which for $(m-1)>k$ it will virtually always have because of random hardware noise in the spectra, and that in practice the full-rank inverse is often replaced by some form of a rank-reduced inverse, i.e., principal component regression (PCR) or partial least squares (PLS). We will not concern ourselves with the type of inverse issue here, but the discussion will return to it later.] The glucose values of the calibration fit are then given by

$$\mathbf{y}_{\text{fit}} = \overline{y_R} + \tilde{\mathbf{X}}\cdot\hat{\mathbf{b}}, \tag{5}$$

and likewise for the future prediction spectra $\mathbf{X}_{\text{pred}}$

$$\mathbf{y}_{\text{pred}} = \overline{y_R} + (\mathbf{X}_{\text{pred}} - \mathbf{1}_{(nx1)}\cdot\overline{\mathbf{x}}^T)\hat{\mathbf{b}}. \tag{6}$$

Everything said so far is good and when applied to a well-designed calibration data set generally produces solutions near the theoretical optimum (described below as the spectrometric Wiener filter). That *optimum cannot be improved upon any further, but the method of arriving at it in practice can be, significantly*, as also can the interpretation and the decision making about where one is in the development process and what to do next.

Historically, the assumption has generally been that the whole error **e** in Eq. (1) is due to inaccuracies in the glucose references $\mathbf{y}_R$ whereas the spectra **X** are assumed to be noise free in most statistics textbooks. This assumption, however, is

invalid in many chemometrical applications where the spectral signal of interest is often buried underneath much larger interfering spectra. We therefore split the calibration spectra into the glucose signal and ''spectral noise:''

$$\mathbf{X}=\mathbf{X}_n+\mathbf{y}\cdot\mathbf{g}^T, \tag{7}$$

where $\mathbf{X}_{n_{(mxk)}}$ is the matrix of spectral noise (AU); $\mathbf{g}_{(kxl)}$ the glucose response spectrum (AU/mg/dL); and $\mathbf{y}_{(mxl)}$ the actual glucose concentrations in the calibration samples (mg/dL). (The response spectrum $\mathbf{g}$ is the spectral signal caused by a change in glucose concentration assuming that everything else stays constant; as such, $\mathbf{g}$ generally depends on the nature of the sample, e.g., gas vs solid, and the nature of the measurement, e.g., transmission vs diffuse reflection.)

After mean centering we have

$$\tilde{\mathbf{X}}=\tilde{\mathbf{X}}_n+\tilde{\mathbf{y}}\cdot\mathbf{g}^T. \tag{8}$$

Here, like in many other applications, only a small part of $\tilde{\mathbf{X}}_n$ is noise from the instrument hardware and most of $\tilde{\mathbf{X}}_n$ is spectral interference from other components in the blood, i.e., water, proteins, etc. From the point of view of glucose measurement, however, spectral noise by definition is everything that is not glucose. The usefulness of this definition will become clear in the following.

Next we need to consider the different types of ''errors'' that can affect the reference values $\mathbf{y}_R$. Our interest is to describe the effect of reference noise on the calibration. Simply defining reference noise as, say, $\mathbf{y}_R-\mathbf{y}$, does not make sense. Assume that the clinical reference method always measured exactly 90% of the true values, i.e., $\mathbf{y}_R=(0.9)\mathbf{y}$. As far as the linear regression is concerned, this would still be perfect reference, although now even a perfect infrared calibration would be biased, $bias=(-)0.1)\overline{y}$, and slope deficient, $slope=0.9$, with respect to the *actual* values. Of course, these errors would not show up on any of the result plots and, in fact, could not be detected unless a second, better, reference method became available. These systematic reference errors are actually quite common in practice, not because of faulty reference analyzers, but because of sample issues. For example, in noninvasive glucose sensing in the skin, the average glucose concentration in the probed tissue volume is lower than the concentration in the blood used for the reference analyses,[15] say, $(\mathrm{mg/dL_{tissue}})=1/2(\mathrm{mg/dL_{blood}})$. Equation (4) will automatically adjust for this kind of internal scaling. [As opposed to another form of scaling, viz., scaling by the user, a trivial example of which is a unit change, e.g., plotting (mmol/L) vs (mg/dL).] Wherever the scaling comes from, the point is we should not call it ''noise.'' Instead, we need to be careful and to clearly divide the responsibilities between the reference method and the infrared measurement.

Bias and slope errors of the reference method with respect to the actual sample concentrations are the mere responsibility of the reference method. From the point of view of the infrared calibration, only the scatter of the reference method, i.e., the part of the vector $\tilde{\mathbf{y}}_R$ that is not correlated with $\tilde{\mathbf{y}}$, can be called reference noise. We therefore split $\tilde{\mathbf{y}}_R$ as follows:

$$\tilde{\mathbf{y}}_R=\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\tilde{\mathbf{y}}_R+\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{y}}_R\equiv S(\tilde{\mathbf{y}}+\tilde{\mathbf{y}}_n), \tag{9}$$

where

$$S\equiv\frac{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}_R}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\quad\text{(the scaling factor between the sample and the reference concentrations), and} \tag{10}$$

$$\tilde{\mathbf{y}}_n\equiv\frac{1}{S}\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{y}}_R\quad\text{(the reference noise vector in mg/dL).} \tag{11}$$

The scaling factor S, like most everything else in calibration, is determined by variances in the concentration signals, and not by their average values. Inserting into Eq. (4) and applying the Sherman–Morrison formula[16] yields

$$\begin{aligned}
\hat{\mathbf{b}}&=[(\tilde{\mathbf{X}}_n^T+\mathbf{g}\cdot\tilde{\mathbf{y}}^T)(\tilde{\mathbf{X}}_n+\tilde{\mathbf{y}}\cdot\mathbf{g}^T)]^{-1}(\tilde{\mathbf{X}}_n^T+\mathbf{g}\cdot\tilde{\mathbf{y}}^T)S(\tilde{\mathbf{y}}+\tilde{\mathbf{y}}_n)\\
&=S\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}_{(mxm)}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n+\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)^T\right]^{-1}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}+\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\\
&=\frac{S\left[\tilde{\mathbf{X}}_n^T\left(I-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})}{1+(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)^T\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)}+\ldots\\
&\quad S\left(\mathbf{I}_{(kxk)}-\frac{\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\cdot\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)^T}{1+(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)^T\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)}\right)\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n\,.
\end{aligned} \tag{12}$$

Equation (12) is the main result of this article and it describes the dependence of the **b** vector on the glucose signal $\tilde{\mathbf{y}}\cdot\mathbf{g}^T$, the spectral noise $\tilde{\mathbf{X}}_n$, the reference concentrations and their noise, $\tilde{\mathbf{y}}_R = S(\tilde{\mathbf{y}}+\tilde{\mathbf{y}}_n)$, and the spurious ''correlations'' $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}$. Notice that the effect of the reference noise on the calibration, in the second summand via $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n$, is usually completely dominated by the effects of the spurious correlations $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}$. Electrical engineers may already recognize the similarities between Eq. (12) and the famous Wiener or ''matched'' filter used in time-signal processing applications, e.g., in cellular phones.

4 Fine-tuning the Theory

Equation (12) looks complicated because it contains all the adverse effects that the user is trying to get rid of in his calibration experiment. If we now make the assumptions that the user has succeeded in sampling a calibration data set in which, first, the effect of the reference noise is zero, $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n = \mathbf{0}$ (this assumption will be made throughout the rest of the article); and second, the effect of spurious correlations is zero, $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}} = \mathbf{0}$. [The discussion will return to spurious correlations later. It can be seen below that spurious correlations can easily be built back in into all formulas, cf., e.g., to Eq. (30), but for simplicity of discussion we throw them out here.] Then Eq. (12) shrinks to

$$\hat{\mathbf{b}} = S\,\frac{(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})}{1+(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\mathbf{g}^T(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}}, \tag{13}$$

which *is* the spectrometric incarnation of the celebrated Wiener filter. That is, the solution, Eq. (13), minimizes, first, the least-squares error of the calibration fit and, second, to the extent that $(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m$ and $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m$ represent the (co-) variances of the future spectral signal and noise, respectively, also the mean-square prediction error of the future spectra.[17]

The Wiener filter is optimal among all **b** vectors in the mean-square error (MSE) sense. Wiener filtering has been extensively used for many decades and in various technical disciplines, mostly time-signal processing. Spectroscopic applications are different from the mainstream in one important point. In time-signal processing, e.g., when detecting the height of an incoming pulse signal, the impulse response (**b** vector) of an electronic Wiener filter is basically determined by the shape of the pulse signal (**g**), because the amplitude of the electronic noise is usually small and its covariance is ''flat,'' i.e., uniform and uncorrelated $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n/m = \sigma^2\mathbf{I}$. On the other hand, in spectrometry the principal reason why people use multivariate techniques is because their pure component signal is buried underneath a large amplitude of spectral noise which, in combination with the fact that the spectral noise is *not* flat, means that the shape of the **b** vector is dominated by the covariance structure (eigenfactors) of the spectral noise and *not* by the signal **g**. This situation has two consequences. First, it makes interpretation of the **b**-vector result itself much harder (see Sec. 6). And, second, in the past it reduced the Wiener filter to an abstract concept rather than a real-world procedure. The fact that Eq. (4) converges against the Wiener filter when the number of calibration samples increases has been repeatedly mentioned in the chemometric literature;[18,19] however, because Eq. (12) was not available, means for the direct insertion of *a priori* physical knowledge about the spectra were not available.

We now proceed by defining the spectral signal-to-noise ratio of the calibration data set as

$$\mathrm{SNR}_x \equiv \sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\mathbf{g}^T(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}}. \tag{14}$$

SNR_x is application specific, i.e., it is different for glucose than it is for, say, cholesterol. SNR_x is very different from the various types of ''hardware SNRs'' that are typically in units of [dc V/root mean square (rms) V] or (dc AU/rms AU), where dc means an average value, and which typically reach values of 10 000 (80 dB) or higher. Values for SNR_x, on the other hand, are much lower. In biomedical applications, $\mathrm{SNR}_x = 10$ is fabulous and many reference methods are just a little above $\mathrm{SNR}_x = 5$. (The main reason for the low values is that the concentrations in the body do not change much to begin with. The important consequences for the development of new biomedical methods will be discussed further below.)

Inserting Eq. (13) for $\hat{\mathbf{b}}$ back into Eq. (5) for the fitted glucose values yields

$$\begin{aligned}\tilde{\mathbf{y}}_{\mathrm{fit}} \equiv \mathbf{y}_{\mathrm{fit}} - \overline{y_R} &= (\tilde{\mathbf{X}}_n + \tilde{\mathbf{y}}\cdot\mathbf{g}^{\mathrm{T}})\cdot S\,\frac{(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})}{1+(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\mathbf{g}^T(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}} \\ &= S\left(\frac{\mathrm{SNR}_x^2}{1+\mathrm{SNR}_x^2}\tilde{\mathbf{y}} + \frac{\tilde{\mathbf{X}}_n(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)^{-1}\mathbf{g}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})}{1+\mathrm{SNR}_x^2}\right),\end{aligned} \tag{15}$$

where the factor $\mathrm{SNR}_x^2/(1+\mathrm{SNR}_x^2)$ in the first term explains the slope deficiency caused by the spectral noise and the second term is the scatter error caused by the spectral noise. The spectral noise $\tilde{\mathbf{X}}_n$, which appears on the right side of Eq. (1), pulls down the magnitude of the **b** vector whereas the reference noise does not. In fact, $\tilde{\mathbf{y}}_n$ does not even appear in Eq. (15) which is a fascinating result in itself because it means that a good *infrared method can be better than the reference method*, i.e., $\|S\,\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_{\mathrm{fit}}\|$ can be smaller than $\|S\,\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_R\|$, provided that the SNR_x is high and a good calibration experiment with $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n = \mathbf{0}$ and $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}} = \mathbf{0}$ is performed. Of course, the fact that the secondary (infrared) method is better than the primary (reference) method cannot be proven unless a second, better reference becomes available.

For the sake of simplified discussion, Eq. (15) has been, and subsequent formulas will be, written for the calibration case and not for the case of independent predictions $\tilde{\mathbf{X}}_{\mathrm{pred}}$. Of course, there is a world of difference between calibration and prediction, but this difference is not amenable to concise mathematical description. The intentions of this article are believed to be better served by the relatively concise formulas derived for the calibration fit because these formulas are probably more helpful in discussing the prediction case than the mathematically more cumbersome formulas containing $\tilde{\mathbf{X}}_{\mathrm{pred}}$. After all, calibration and prediction results are the same in the

probability limit (meaning, for *large* numbers of calibration and prediction spectra) *if* the calibration and prediction spectra are measured in identical ways (meaning, if the calibration and prediction spectra come from the same underlying distribution). In practice, there are many reasons why the calibration and prediction spectra may not be measured in identical ways but the goal is that they are. The most important reason for differences in practice is probably the potential for long-term drift in the instrument and/or sample, causing a ''slowly increasing bias'' in the predictions. (Notice that the distinction between bias error and scatter error is a purely practical matter and that a mayfly and a Galapagos turtle would have different opinions on the subject.) Bias will not get much coverage in this article, however, it is important to realize that, in ill-posed systems in particular, control of bias is a problem second to none and often constitutes the ultimate engineering challenge. To repeat, long-term independent prediction performance is the goal and here we will use the calibration case only as a vehicle to talk about this goal.

Two more definitions are needed. The signal-to-noise ratio of the reference method is

$$\mathrm{SNR}_y \equiv \sqrt{\frac{\tilde{\mathbf{y}}^T \cdot \tilde{\mathbf{y}}}{\tilde{\mathbf{y}}_n^T \cdot \tilde{\mathbf{y}}_n}}, \tag{16}$$

and the total SNR of the calibration data set is

$$\mathrm{SNR} \equiv \sqrt{\frac{\mathrm{SNR}_x^2\,\mathrm{SNR}_y^2}{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2}}. \tag{17}$$

Now, then, when we measure the slope what we do is plot $(\tilde{\mathbf{y}}_{\mathrm{fit}}+\tilde{\mathbf{y}}_R)$ vs $(\tilde{\mathbf{y}}_R+\tilde{\mathbf{y}}_R)$ and LS fit a straight line through the scatter cloud. Here, the reference noise $\tilde{\mathbf{y}}_n$ enters back into the picture because, even though the **b** vector is not affected by $\tilde{\mathbf{y}}_n$ by virtue of the assumed $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n=\mathbf{0}$, after some lengthy algebra the measured slope comes out to be

$$\mathrm{slope} \equiv \frac{\tilde{\mathbf{y}}_{\mathrm{fit}}^T\tilde{\mathbf{y}}_R}{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R} = \frac{\mathrm{SNR}_x^2}{1+\mathrm{SNR}_x^2}\,\frac{\mathrm{SNR}_y^2}{1+\mathrm{SNR}_y^2} = \frac{\mathrm{SNR}^2}{1+\mathrm{SNR}^2}. \tag{18}$$

So the slope is pulled down twice. First, the spectral noise $\tilde{\mathbf{X}}_n$ pulls down the **b** vector and thereby the predictions and, second, the reference noise $\tilde{\mathbf{y}}_n$ decreases the slope at the time at which the line is fitted in the scatter plot.

The correlation coefficient between the predicted and the reference concentrations is

$$r \equiv \frac{\tilde{\mathbf{y}}_{\mathrm{fit}}^T\tilde{\mathbf{y}}_R}{\sqrt{(\tilde{\mathbf{y}}_{\mathrm{fit}}^T\tilde{\mathbf{y}}_{\mathrm{fit}})(\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R)}} = \sqrt{\frac{\mathrm{SNR}_x^2}{1+\mathrm{SNR}_x^2}\,\frac{\mathrm{SNR}_y^2}{1+\mathrm{SNR}_y^2}} = \sqrt{\frac{\mathrm{SNR}^2}{1+\mathrm{SNR}^2}}. \tag{19}$$

The infrared method is responsible for SNR_x, the reference method is responsible for SNR_y, and the calibration is left to cope with both, SNR. Equation (19) corresponds to the expected $r^2=r_x^2r_y^2$ where $r_x^2\equiv\mathrm{SNR}_x^2/(1+\mathrm{SNR}_x^2)$ and the same for r_y^2. Correlation coefficient and SNR are synonymous concepts that measure the same thing. The use of SNR is preferred by this author, however, because it is easier to interpret. One can ''feel'' the huge difference among SNR=1, 10, and 100 whereas $r=0.71$, 0.995, and 0.99995 is much harder to interpret. Also, the term ''signal-to-noise ratio'' is a constant reminder of the fact that the quality of a calibration depends on both the signal and noise, and that simple comparisons between calibrations are fair only at identical signal levels $\sqrt{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}$.

The slope error is

$$\begin{aligned}\text{slope error} &\equiv \sqrt{(\tilde{\mathbf{y}}_R-\text{slope }\tilde{\mathbf{y}}_R)^T(\tilde{\mathbf{y}}_R-\text{slope }\tilde{\mathbf{y}}_R)}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2}{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2+\mathrm{SNR}_x^2\,\mathrm{SNR}_y^2}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{1}{1+\mathrm{SNR}^2}.\end{aligned} \tag{20}$$

Slope error is always present but becomes evident only when the SNR is worse than about 5. In the extreme case in which the SNR approaches 0, the slope has to turn to zero because the best the Wiener filter is left to do is to predict the flat average, $\overline{y_R}$.

The scatter error is

$$\begin{aligned}\text{scatter error} &\equiv \sqrt{(\tilde{\mathbf{y}}_{\mathrm{fit}}-\text{slope }\tilde{\mathbf{y}}_R)^T(\tilde{\mathbf{y}}_{\mathrm{fit}}-\text{slope }\tilde{\mathbf{y}}_R)}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{\mathrm{SNR}_x}{1+\mathrm{SNR}_x^2}\,\frac{\mathrm{SNR}_y}{1+\mathrm{SNR}_y^2}\\ &\quad\times\sqrt{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{\mathrm{SNR}}{1+\mathrm{SNR}^2}.\end{aligned} \tag{21}$$

The total prediction error, aka $\mathrm{PRESS}^{1/2}$, is (assuming zero bias)

$$\begin{aligned}\mathrm{PRESS}^{1/2} &= \sqrt{(\tilde{\mathbf{y}}_{\mathrm{fit}}-\tilde{\mathbf{y}}_R)^T(\tilde{\mathbf{y}}_{\mathrm{fit}}-\tilde{\mathbf{y}}_R)}\\ &= \sqrt{\text{slope error}^2+\text{scatter error}^2}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\sqrt{\frac{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2}{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2+\mathrm{SNR}_x^2\,\mathrm{SNR}_y^2}}\\ &= \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{1}{\sqrt{1+\mathrm{SNR}^2}},\end{aligned} \tag{22}$$

and is the minimum RSS prediction error that can be achieved for a given SNR, i.e., it is the Wiener filter result. The Wiener filter achieves its optimality by trading off scatter versus slope error in a RSS-optimum way. In some applications, however, it is reasonable to require a prediction slope of one even when the SNR is low, because the price paid in increased $\mathrm{PRESS}^{1/2}$ is considered worth the improved accuracy when measuring at the low and high ends of the concentration range. In situations in which slope compensation is an issue, the **b** vector can simply be multiplied by a scalar>1, which can be defined at the user's discretion. The multiplication will, of course, disturb the optimality of the Wiener filter; e.g., the ''100%-slope-corrected'' prediction error,

$$\mathrm{PRESS}^{1/2}_{\mathrm{slope}=1} \equiv \sqrt{\left(\frac{\tilde{\mathbf{y}}_{\mathrm{fit}}}{\mathrm{slope}}-\tilde{\mathbf{y}}_R\right)^T\left(\frac{\tilde{\mathbf{y}}_{\mathrm{fit}}}{\mathrm{slope}}-\tilde{\mathbf{y}}_R\right)} = \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{\sqrt{1+\mathrm{SNR}_x^2+\mathrm{SNR}_y^2}}{\mathrm{SNR}_x\mathrm{SNR}_y} = \sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}\,\frac{1}{\mathrm{SNR}}, \tag{23}$$

is larger than the Wiener filter result. Equation (23) is an intuitive result, saying that the total noise in the calibration data set is the scatter around the 100%-slope-corrected, aka identity, line. Of course, there is no such thing as a ''correct'' slope. The physics of the pure component spectra and the spectral noise are manifested in the *shape* of the **b** vector, not in its magnitude. By default, so to speak, the slope is at the Wiener value, but, in general, users are free to find their own best trade-off between the mean-square prediction error and end-of-range accuracy. In the following, we will therefore apply the term ''Wiener filter'' loosely to both the original and its slope-corrected versions. Fortunately, slope correction becomes an issue only when the SNR<5.

Because the user is free to correct for slope deficiency at his own discretion, $\mathrm{PRESS}^{1/2}$ is not a unique measure of calibration quality. On the other hand, the correlation coefficient and SNR are unique measures of calibration quality because the user changing the slope does not affect them.

The total SNR of a given data set can be measured in a number of ways using Eqs. (18)–(23). Which one to use is a matter of convenience and depends on the situation, however, some caution is advised because some situations are tricky. For example, the calibration spectra are often more extensively averaged than the later prediction spectra. This is commonly done to reduce spectral noise in the calibration data set and to trick the Wiener filter into producing higher slopes. In this situation, where the calibration SNR is actually different from the prediction SNR, a reasonable choice might be to use the prediction slope to measure the calibration SNR and to use the prediction correlation coefficient to measure the prediction SNR.

Unfortunately, SNR_x and SNR_y cannot be *individually* determined from a single calibration experiment because the calibration is only affected by their combined total, SNR. (By the way, exchanging the x and y axes of the scatter plot produces $\mathrm{slope}_{x-vs-y}\equiv 1$.) For many applications, multiple calibration experiments do not help either because SNR_x and SNR_y scale identically with the signal. One trick that can be used to overcome this situation is to perform two calibrations, with different signal levels, and to intentionally degrade the SNR_y of one. This is, in fact, exactly what happens with many of the ''wet chemical'' reference methods anyway, which are typically dominated by multiplicative errors, i.e., small concentrations are measured with small errors and large concentrations are measured with large errors. Assume that two calibration data sets have been collected under virtually identical spectroscopic conditions, that one happens to have significantly more signal $\sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R/m}$ than the other, and that still the two SNRs come out to be the same. The typical explanation for this is that SNR$\cong\mathrm{SNR}_y<\mathrm{SNR}_x$ and SNR_y $=\mathrm{constant}(\sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R})$.

The results of Sec. 4 are summarized in Figure 2, which shows the other statistics to be highly nonlinear functions of the SNR that suddenly ''come out the noise'' at and above SNR$=1$. The Monte-Carlo generated example scatter plots in Figure 3 demonstrate the rapid and quite dramatic improvement in the visual appearance of a scatter plot once the SNR improves to above 1. The range from approximately SNR $=(0.5)$ to 2 is called the ''cliff'' by this author. Operating in this region is a tiring experience for the technical staff in many companies.

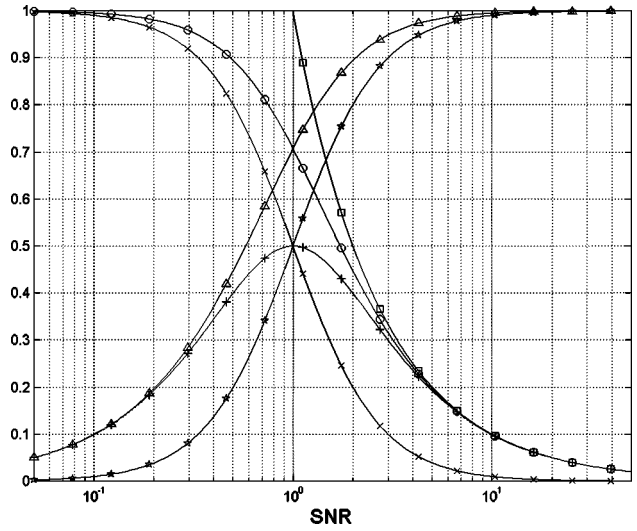

Fig. 2 Calibration statistics as a function of the SNR: prediction slope (☆), correlation coefficient (△), slope error (×), scatter error (+), $\mathrm{PRESS}^{1/2}=\sqrt{\text{slope error}^2+\text{scatter error}^2}$ (○), and slope corrected $\mathrm{PRESS}^{1/2}_{\mathrm{slope}=1}$ (□). The errors are all normalized by dividing by $\sqrt{\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R}$.

5 Discussion

A number of practically important issues that typically come up in the practice of applying multivariate calibration techniques to spectroscopic data are discussed next. It is hoped

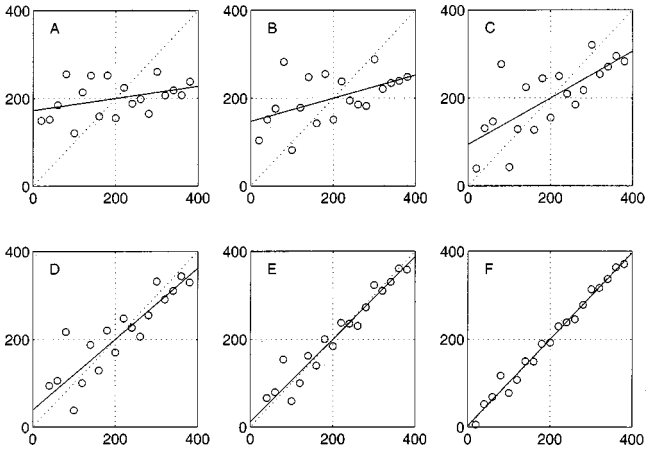

Fig. 3 Monte-Carlo generated example demonstrating the strong effect of the SNR on the visual appearance of the scatter plot: SNR = (A) 0.25, (B) 0.5, (C) 1, (D) 2, (E) 4, and (F) 8. *A posteriori* LS-fitted lines (solid), identity lines (dashed). Bias and reference noise set to zero in this simlation.

that here in Sec. 5 elucidating many of the consequences, and opportunities, hidden in the mathematical sections above will help.

5.1 Physics Behind the "Statistical" Model

The Wiener filter solution, Eq. (13), of the statistical model, Eq. (1), makes perfect physical sense. To illustrate in a first example, assume that the scaling factor is $S=1$ and the rms spectral noise is flat, i.e., uniform on all pixels and not correlated between any two pixels $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m \equiv \sigma_x^2\mathbf{I}_{(kxk)}$ (AU^2). Then define the rms signal as $(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m = s_y^2$ $(\mathrm{mg/dL})^2$. Equation (13) then reads

$$\hat{\mathbf{b}} = \frac{s_y^2}{\sigma_x^2}\frac{\mathbf{g}}{1+s_y^2(\mathbf{g}^T\sigma_x^{-2}\mathbf{g})} \quad [(\mathrm{mg/dL})/\mathrm{AU}], \tag{24}$$

where the shape of the **b** vector is *identical* to the shape of the pure glucose response spectrum. As a second example, consider the limit of disappearing spectral noise. Using Eq. (24) and letting $\sigma_x^2 \to 0$, we have the ideal,

$$\hat{\mathbf{b}} \to \frac{\mathbf{g}}{\mathbf{g}^T\mathbf{g}} \quad [(\mathrm{mg/dL})/\mathrm{AU}]. \tag{25}$$

The reason that the shape of real **b** vectors does not look like the infrared glucose spectrum is that the real spectral noise is not uniform and is correlated between pixels. To conclude that "statistical" models are not "physical" is wrong. In fact, there is no fundamental difference at all because both approaches even follow the same basic idea, viz., to find the direction of maximum SNR_x. Consider a third example, where two wavelengths are used to measure an "analytical" absorption band at wavelength λ_1 in the presence of large baseline variations. Here, wavelength λ_2 is selected next to the absorption band and the spectral noise is $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m = \sigma_x^2\left(\begin{smallmatrix}1 & 1\\ 1 & 1\end{smallmatrix}\right)$ and the response spectrum is $\mathbf{g}=\left(\begin{smallmatrix}g\\ 0\end{smallmatrix}\right)$. Inserting into Eq. (13) yields the Wiener filter

$$\hat{\mathbf{b}} = \frac{\frac{1}{4}\left(\frac{s_y}{\sigma_x}\right)^2 g}{1+\frac{1}{4}\left(\frac{s_y}{\sigma_x}\right)^2 g^2}\begin{pmatrix}+1\\ -1\end{pmatrix}, \tag{26}$$

which, except for the slope correction, is identical to the expected physical result. Both statistical and physical models try to point their **b** vectors away from the direction of maximum signal (**g**) and into the direction of maximum SNR_x. The basic difference is that the statistical models use an actual measurement of the spectral noise, $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n$, as presented in the calibration data set, whereas the physical models rely on human intuition to describe the spectral noise. The two produce virtually identical results in simple cases. When it comes to spectroscopy of complex samples, e.g., near-IR, spectroscopy, however, human intuition can no longer compete.

5.2 Is There a Better Way for Multivariate Calibration?

Yes, one can help the statistical model converge against the Wiener filter faster. The ways in which to insert *a priori* knowledge are numerous and application specific, but the core statement of this article is this: One can combine different pieces of *a priori* physical knowledge about the spectra with any available measured data to estimate the pure-component spectral signal and spectral noise separately, and then compute the Wiener filter "manually" by plugging the results into Eq. (13) (see Sec. 6). The effects of spurious correlations and reference noise are eliminated right off the bat and the quality of the estimate of the Wiener filter is limited only by the quality of the initial estimates of the spectral signal and noise. Specificity can be guaranteed because the spectral signature of the signal (**g**) is under user control. Important trade-off decisions concerning calibration transfer or long-term stability can be made by adjusting the estimate of the spectral noise (e.g., whether or not to include instrument-to-instrument noise). In a fortunate case in which both **g** and $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m$ are known, collection of further calibration spectra is not necessary at all and users can directly compute the Wiener filter [and if desired slope adjust it according to the expected $(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m$ in the application]. In a more typical case, spectral noise is not known and calibration samples will still have to be collected to estimate the spectral noise, however, *reference analyses are not necessary as soon as the slope of* **g** *is known*.

When **g** is not known, then there are still multiple ways by which to improve the quality of the estimate but now they are not as obvious (e.g., one can incorporate *a priori* knowledge about the spectral regions in which **g** does *not* have any bands). We defer to later publications, however, we do want to mention that cases in which a good traditional calibration is available as a "starting ground" are especially fruitful to work with.

We also mention in passing that, mathematically, **g** does not have to be a *pure* component response spectrum to apply Wiener filtering. Any mixture spectrum (whose scaled mixture concentration might be of interest for some reason) could be used as well. This situation is more common in process-control applications; e.g., when measuring the concentration of latex in paper coatings, the response spectrum of the latex depends on its composition and thus on the manufacturer.

5.3 Is There an Optimum Wavelength Range?

No, mathematically there is no "optimum" λ range because wider is always better/equal. This is because $\tilde{\mathbf{X}}_n^T\mathbf{X}_n$ is a positive *semi* definite matrix, and SNR_x will therefore always increase or at least stay equal when the number of channels is increased. (The same reason why calibration results always get better/equal with increasing numbers of λ channels or PLS/PCR ranks.) So the practical challenge is to find the most SNR_x bang for the least hardware bucks. Equation (14) can be favourably used in the search of a "good" subset of wavelength channels.

But there is also the basic limitation hidden in the word "semi" above, viz., the limited information content of the spectrum itself. Additional λ channels can only improve the SNR_x if the added λ channels either contain new glucose features, or contain spectral noise that is correlated with the (still uncorrelated) spectral noise in the λ areas of the "old" glucose features, or both. This theoretical limit can be relevant in practice. [In the example, Eq. (26), we saw how the SNR_x can be improved by including λ regions with zero glucose signal, viz., by subtracting out noises that are correlated between pixels.]

Spectral noise $\tilde{\mathbf{X}}_n$ generally consists of three independent parts: *sample noise*, *sampling noise*, and *instrument noise*. Sample noise is from spectral interference by the other components in the sample, sampling noise is from variations in the sample handling, and instrument noise is from the hardware. Whereas the first two are typically correlated between pixels (broad spectral features), the last one is typically not. Since it cannot be subtracted out between pixels, the signal vector $\mathbf{g}\sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m}$ must peak out over the rms instrument noise floor (AU) if the calibration is to stand any chance. Adding more and more pixels that do not contain glucose signals does not help the instrument noise situation.

5.4 Orthogonalization into Many Univariate Regressions and Visualization

With regard to the statement that ''spectrum interpretation is only an afterthought in the NIR,'' [20] unfortunately, this is true to a large extent. Probably the best way to tackle the visualization problem is to think of the multidimensional regression problem as a multitude of one-dimensional regression problems by using the singular value decomposition of spectral noise,

$$\tilde{\mathbf{X}}_n = \mathbf{U}\cdot\mathbf{S}\cdot\mathbf{V}^T = (\mathbf{u}_1 \;\; \mathbf{u}_2,\ldots,\mathbf{u}_k)\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_k \end{bmatrix} \times[\mathbf{v}_1 \;\; \mathbf{v}_2,\ldots,\mathbf{v}_k]^T, \tag{27}$$

as the coordinate system. Here, the $\mathbf{u}_i$ and $\mathbf{v}_i$ are the eigenvectors in the time and spectral domains, respectively, ranked in order of $s_1 \geqslant s_2 \geqslant \ldots \geqslant s_k$. By defining $\mathbf{g}_\mathbf{v} \equiv \mathbf{V}^T\mathbf{g}$ and $\hat{\mathbf{b}}_\mathbf{v} \equiv \mathbf{V}^T\hat{\mathbf{b}}$, Eqs. (13) and (14) can be written in orthogonal form,

$$\hat{\mathbf{b}}_\mathbf{v} = S\frac{\mathbf{S}^{-2}\mathbf{g}_\mathbf{v}(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})}{1+(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\mathbf{g}_\mathbf{v}^T\mathbf{S}^{-2}\mathbf{g}_\mathbf{v}}, \tag{28}$$

and

$$\mathrm{SNR}_x = \sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\mathbf{g}_\mathbf{v}^T\mathbf{S}^{-2}\mathbf{g}_\mathbf{v}} = \sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\left(\frac{g_{v1}^2}{s_1^2}+\frac{g_{v2}^2}{s_2^2}+\ldots+\frac{g_{vk}^2}{s_k^2}\right)}, \tag{29}$$

where $\mathrm{SNR}_{x,i} \equiv \sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})(g_{vi}^2/s_i^2)}$ is the SNR_x in the direction of the ith eigenvector. The infrared measurement of blood glucose is called ''ill posed'' because the glucose signal $\sqrt{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\cdot\|\mathbf{g}\|$ is smaller than many of the larger singular values s_i of spectral noise. This means, first, that the SNR_x found in the data must come from the ''smaller'' eigenfactors, and, second, that the naked eye cannot see any glucose features in the spectra. (In this article, the terms larger and smaller eigenfactors are used to refer to the eigenfactors of spectral noise with larger and smaller eigenvalues, respectively.) In practice, some visualization can be recovered by using the following simple procedure. Compute the SVD of the measured $\tilde{\mathbf{X}} \equiv \tilde{\mathbf{X}}_n + \tilde{\mathbf{y}}\cdot\mathbf{g}^T$; then compute the correlation coefficients of the resultant time eigenvectors with the glucose reference concentrations $\tilde{\mathbf{y}}_R$. The eigenfactors with the largest correlation coefficients will dominate the prediction results and their spectral eigenvectors should resemble the glucose spectrum (if the calibration is not seriously affected by spurious correlations). The resemblance will be modest because of the algebraic constraints on the eigenvectors, but the peaks will be at the right place and will have the right sign and magnitude. Since the $\mathrm{SNR}_{x,i}$'s in Eq. (29) add up in squares the resulting SNR_x is usually dominated by only a few spectral directions.

We point out here that a multivariate measurement can be ill posed regardless of whether or not the associated matrix inversion is ''ill conditioned.'' The two concepts are different. Statisticians interested in parameter estimation use the condition number but this measure is largely irrelevant in chemometrics where the goal is prediction. For example, a calibration matrix made up of near-IR spectra of liquid samples measured at $\Delta\lambda = 0.1$ nm spectral resolution will be very ill conditioned but just as well posed, or even slightly better posed, than a matrix containing the same samples measured at, say, $\Delta\lambda = 10$ nm.

5.5 Small Spectral Signals Can Still Have High SNR_x (and Vice Versa)

The concepts of SNR_x and ill posedness go hand in hand but are *not* identical. Theoretically, e.g., the 20th eigenfactor of a noninvasive glucose measurement system could carry a very high $\mathrm{SNR}_{x,i} = 10$ (just an example); this system would be able to predict with a perfect $\mathrm{SNR}_x \geqslant 10$ yet still it would be very ill posed, because the calibration would need to ''dig away'' 19 larger eigenfactors of spectral noise to get to the SNR_x. In reality, near-IR noninvasive glucose sensing belongs to the tough class of problems that are both ill posed and have low SNR_x. As far as the hardware is concerned, the engineering has to solve two problems: the system noise must be reduced to the point where some of the smaller eigenfactors can deliver the needed SNR_x and spectral noise in the larger eigenfactors must be prevented from ever spilling down into the smaller, high-SNR_x eigenfactors. The capability to do the latter is one of the important characteristics that distinguishes a good piece of hardware.

5.6 Number of Calibration Samples

The important question of how many independent calibration samples are needed is governed by many practical issues, including spurious correlations/overfitting, quality of the statistical estimate of the spectral noise, and ''riding the cliff.'' These issues are discussed next. We point out that formal statistical tests can also be performed (e.g., based on those in Ref. 21) but here we will focus on the more practical aspects instead.

5.7 Spurious Correlations/Overfitting

As discussed above, if the spectral signal $\mathbf{g}$ is known, then the best way to eliminate spurious correlations is to have the user himself define what the spectral signal is and what the spectral noise is and then insert these estimates into Eq. (13). Any physical *a priori* knowledge available about the spectral signal and spectral noise can be combined with any measurements available, and used directly to estimate the optimum Wiener $\mathbf{b}$ vector. The danger of spurious correlations is completely avoided. An example of this ''direct'' way of estimating the Wiener filter will be given below. Still, spurious cor-

relations continue to be a challenge in situations in which **g** is not known and this case will be discussed next.

Equation (12) says that, if $\|\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}/(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})\| \ll \|\mathbf{g}\|$, then the calibration is guaranteed not to be affected. This indicates a catch-22 situation: in order to *prove* that a calibration is not affected, **g** must be known. In the past, what was sometimes done to prove specificity *a posteriori* was to plot the measured $\tilde{\mathbf{X}}^T\tilde{\mathbf{y}}_R$ (or a scaled version of it called the property-correlation spectrum[22]) and compare it to the shape of the known glucose spectrum. If the two looked alike, then the calibration was judged good. (Mathematically, this procedure is not completely correct because the spurious correlation spectrum could happen to be exactly parallel to **g**; however, apart from being very unlikely, this result would not change the shape of the **b** vector, only its magnitude, which is subject to slope correction by the user anyway. In practice, what is feared about spurious correlations is changes to the *shape* of the **b** vector, not in its magnitude.) In the case of infrared blood glucose analysis and typical-size calibration data sets ($m = 100...300$), this ''visual test'' typically ended positive on mid-IR spectra but generally negative on near-IR spectra. What was the conclusion when the test failed? None, inconclusive. The visual test is sufficient but is not necessary. The necessary and sufficient condition for spurious correlations to be negligible is that the $\mathrm{SNR}_{x_{sp}}$ with spurious correlations is only insignificantly larger than the SNR_x from the true signal alone or, mathematically speaking,

$$\mathrm{SNR}_{x_{sp}} \equiv \sqrt{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)^T\left[\tilde{\mathbf{X}}_n^T\left(\mathbf{I}-\frac{\tilde{\mathbf{y}}\cdot\tilde{\mathbf{y}}^T}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)\tilde{\mathbf{X}}_n\right]^{-1}\left(\mathbf{g}+\frac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}\right)} \cong \sqrt{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}\cdot\mathbf{g}^T(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_\mathbf{n})^{-1}\mathbf{g}} = \mathrm{SNR}_x. \quad (30)$$

What matters to calibration is SNR_x, which is equivalent to correlation, not covariance. For example, given a data set with a tiny glucose signal of, say, 10 μAU rms with SNR = 2 and a huge humidity effect of 10 mAU rms with SNR = 0.2, the **b** vector will still lock onto the glucose information almost exclusively, leaving the predictions virtually unaffected by humidity. An example of this behavior will be given below.

Equation (30) is a nice piece of background information but, when **g** is not known, it does not provide a practical way in which to deal with spurious correlations. Therefore, if no *a priori* information about the spectral signal **g** is available, then the only reliable way to safeguard against spurious correlations is to perform extensive randomized calibration experiments in the ''traditional'' way, i.e., via Eq. (4). In practice, the proof of the method then comes gradually over time when multiple such randomized experiments performed in a development program consistently deliver identical looking **b** vectors, with the SNR coming from the same spectral eigenvectors.

The core of the spurious correlation problem with traditional calibrations usually is that some of the larger time eigenvectors of the spectral noise are not rapidly fluctuating aka ''random'' functions of time but are slowly undulating drifts (slow on a human scale). If the characteristic time constant of a slow process is, say, 3 h, then independent samples can only be measured at >3 h intervals (Nyquist's sampling theorem). Slow spectral noises therefore need extra attention to decorrelate them from $\tilde{\mathbf{y}}_R$. The best way is to randomize $\tilde{\mathbf{y}}_R$ and this is standard procedure in virtually all *in vitro* experiments. In some applications, e.g., noninvasive glucose sensing, effective randomization requires long calibration time periods on the order of several weeks. The minimum time required can be estimated as follows. Assume the true-signal SNR_x is 2. In order to minimize the effect of spurious signals, the false-SNR_x from spurious signal alone is required to be smaller than, say, 0.4. Say that there are five slow time processes in the spectral noise that each could correlate spuriously. Say we require that each process only correlates with SNR_x of $0.4/\sqrt{5} = 0.179$ which is equivalent to $r = 0.176$. In statistics books[23] it is said that in order to achieve $|r| < 0.176$ between two sets of random numbers with 95% probability, the number of random pairs needs to be larger than ~120. Thus, the calibration experiment should collect at least 120 *independent* calibration samples.

In the example above, 120 samples were enough to break the spurious correlations to the slow spectral processes, which typically reside in the larger eigenfactors, but 120 may or may not be enough to also diminish the effect of *overfitting*. This term is loosely used in the chemometrical literature to describe spurious correlations in the *smaller*, noisy-looking eigenfactors (as opposed to the statistical literature, where the same term is used to describe the inclusion of too many variables in a statistical model). The spurious correlation in each small eigenfactor may be small, but many of them can add up. A standard rule of thumb used in statistics to control overfitting is to use at least five or six times as many samples as variables, and the same rule is also recommended as standard practice for chemometrics (where variables are defined as either wavelengths or PLS/PCR factors).[24] Following this rule will actually do two things. First, spurious correlations to the instrument noise aka overfitting is reduced and, second, the quality of the statistical estimate of the covariance matrix of spectral noise is improved (see Sec. 5.8). A practical way to check for overfitting is described in connection with Figure 8(b). If overfitting is a problem, then PLS or PCR can be used advantageously to cut out affected eigenfactors. PLS or PCR should not be relied on excessively in this regard, however, because the appearance of spurious correlations in the smaller eigenfactors usually also indicates bad quality of the estimate of the true spectral noise, in which case the only way to proceed is to increase the number of calibration samples.

5.8 Quality of the Statistical Estimate of Spectral Noise

In the special case of the true covariance matrix being from uniform, uncorrelated noise, i.e. $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m \to \sigma_x^2\mathbf{I}_{(kxk)}$ for $m \to \infty$, there is a simple graphical ''eigenvalue flatness test'' of the quality of the statistical estimate. Plotting the eigenvalues of $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m$ for a finite number of samples m yields a sloped trace of eigenvalues, instead of the ideal flat one. Thus, the higher the number of calibration samples, the flatter the eigenvalue trace of the instrument noise floor sampled, which is a graphical expression of the 5× rule mentioned above. The flatness test can be used as a practical guideline for the number of calibration samples required in real data sets. Using MATLAB notation, try `plot(1:k,svd(randn(m,k))/sqrt(m))` with different values of m to find your own tolerance level for flatness. Fortunately, flatness of the *smaller* eigenvalues can

often be improved by repeated measurements. For example, in the previous example, performing 120 measurement sessions with three repeats each will yield 360 independent measurements of the fast instrument noises in the smaller eigenvalues whereas, of course, there will still be only 120 independent realizations of the slow spectral processes in the larger eigenvalues. A more accurate description of how many independent samples it takes to estimate the eigenfactors of a particular multidimensional measurement system in a statistically reliable way is given in Anderson's theorem.[25]

5.9 Cross-Validation

The issue of cross-validation, e.g., ''leave-one-out'' cross-validation, is closely related to the issue of spurious correlation. Whether the cross-validation results are useful or useless, i.e., do or do not resemble independent predictions, depends on whether or not the reduced-subset calibrations are affected by spurious correlations and whether or not the left out and predicted spectra can take direct advantage of these spurious correlations. By far the most notorious example of cross-validation run amok is the oral glucose tolerance test (OGTT) used for noninvasive blood glucose calibration. In such an OGTT, a diabetic patient drinks sugar syrup causing his glucose concentration to go up and, after insulin injection, back down again. The whole exercise may last 8 h and may result in hundreds of skin spectra (infrared or other) collected during that time period. What we have is (i) a multitude of slow spectral processes well above the instrument noise floor sampled with fewer than 10 independent samples (Nyquist); (ii) fewer than 10 independent samples of $\tilde{\mathbf{y}}_R$ (Nyquist); and (iii) left out and predicted spectra that can take full advantage of any spurious correlations. Single-day OGTT cross-validation will produce nice looking scatter plots that ''predict'' virtually any $\tilde{\mathbf{y}}_R$ time profile under the sun. In the hands of an inexperienced user and without the background of a prior good calibration experiment, OGTT results are generally worthless. On the other hand, and concluding this discussion with one positive remark about cross-validation, for experiments in which the sequence of samples is fully randomized, e.g., in many *in vitro* studies, cross-validation results can be close to truly independent prediction results, with the exception of bias, of course.

5.10 Riding the Cliff

There is another effect, called ''riding the cliff'' by this author, which is not as well known as spurious correlations but is the second most-frequent reason for wild goose chase R&D efforts. Riding the cliff occurs whenever the true SNR is in the region of the cliff (cf. Figure 2) and small changes in the calibration SNR—due to the large appertaining changes in the visual appearance of the scatter plots—trigger a series of wrong conclusions, always one step behind the latest result. Anything that affects the sampled calibration SNR can cause this effect, e.g., spurious correlations, but it is often overlooked how easily the effect can be set off even in seemingly ''innocent'' situations. Glucose is not a good example here, and we will use hemoglobin instead. Hemoglobin is a typical biomedical analyte that varies very little between patients and within a patient over time. The normal physiological range is from about 12 to 16 g/dL yielding signals of about 1 g/dL rms around an average of 14 g/dL. Assume that the calibration signal can be increased to, say, 1.5 g/dL rms by selecting more patients from the ends of the concentration range and that the development effort has achieved a promising SNR = 1.5 at that signal level. Further assume that the development process works by conducting a series of randomized calibration experiments with, say, 20 patients each. [Hemoglobin has large absorbance signals in the visible so systems can use very few wavelengths.] Even if spurious correlations are assumed to be negligible, the laws of statistics will still work against the company because of the low number of calibration samples. Say that the signal level can be reproduced at that value of 1.5 g/dL rms from experiment to experiment. The standard deviation of noise, however, will vary 1 ± 0.32 g/dL rms just by random chance (95% confidence limit) meaning that the SNR sampled in any experiment can vary anywhere from 1.1 to 2.2, causing dramatic differences in the appearance of the scatter plot. (Above, we used the rule of thumb that the standard deviation of the standard deviation is $100\%/\sqrt{2m}$ of the true standard deviation, where m is the number of the independent samples.) A lot of management decisions, including PR and HR decisions, can be as random as the noise that caused them. It should be noted here that the current standard[24] calls for a minimum of 24 calibration samples and the point here is that 24 is too low when the SNR of the application is in the cliff.

5.11 Unspecific Correlations

A very important issue is what this author calls ''unspecific correlations,'' as opposed to spurious ones. Mathematically, the two can be considered identical, but practically they *are* different. Whereas spurious correlations change randomly from experiment to experiment, unspecific correlations are physically unspecific but statistically *reproducible*. Again, glucose is not a good example here, and we will use albumin instead. Imagine the task of calibrating an *in vitro* IR spectroscopic blood analyzer to albumin. Albumin does not vary much within a patient over time so the rms calibration signal has to come from patient-to-patient variations in the calibration set. However, the patient-to-patient variation of albumin correlates well with that of total protein ($r > 0.9$), and this correlation is statistically reproducible from data set to data set. In the traditional way of statistical calibration, the algorithm is therefore never told to only use albumin's pure component spectrum as a ''signal,'' and to shrink the **b** vector in the subspace affected by the other proteins' absorbance features because they are ''noise;'' instead, the traditional solution utilizes absorbance features from the other proteins to predict albumin. We quantify this by rewriting Eq. (8) as $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_{nn} + \tilde{\mathbf{y}}_2\mathbf{g}_2^T) + \tilde{\mathbf{y}}_1\mathbf{g}_1^T$, where $\tilde{\mathbf{y}}_1\mathbf{g}_1^T$ is the spectral signal of the analyte of interest (albumin) and the term in parenthesis is the spectral noise, now consisting of $\tilde{\mathbf{y}}_2\mathbf{g}_2^T$ (the sum of the other proteins) and $\tilde{\mathbf{X}}_{nn}$ (the spectral noise from all other things). We define the correlation coefficient $r_{12} = \tilde{\mathbf{y}}_2^T\tilde{\mathbf{y}}_1 / \sqrt{(\tilde{\mathbf{y}}_1^T\tilde{\mathbf{y}}_1)(\tilde{\mathbf{y}}_2^T\tilde{\mathbf{y}}_2)}$ and assume that the spurious correlations are zero, i.e., $\tilde{\mathbf{X}}_{nn}^T\tilde{\mathbf{y}}_1 = \mathbf{0}$ and $\tilde{\mathbf{X}}_{nn}^T\tilde{\mathbf{y}}_2 = \mathbf{0}$, and that the effects from the reference noise are zero, i.e., $\tilde{\mathbf{X}}_{nn}^T\tilde{\mathbf{y}}_{n1} = \mathbf{0}$ and $(\tilde{\mathbf{y}}_2\mathbf{g}_2^T)^T\tilde{\mathbf{y}}_{n1} = \mathbf{0}$. Inserting into Eq. (12),

$$\hat{\mathbf{b}}=\frac{S\,[\tilde{\mathbf{X}}_{nn}^{T}\tilde{\mathbf{X}}_{nn}+(\tilde{\mathbf{y}}_{2}^{T}\tilde{\mathbf{y}}_{2})(1-r_{12}^{2})\mathbf{g}_{2}\mathbf{g}_{2}^{T}]^{-1}\left(\mathbf{g}_{1}+\mathbf{g}_{2}r_{12}\sqrt{\frac{\tilde{\mathbf{y}}_{2}^{T}\tilde{\mathbf{y}}_{2}}{\tilde{\mathbf{y}}_{1}^{T}\tilde{\mathbf{y}}_{1}}}\right)(\tilde{\mathbf{y}}_{1}^{T}\tilde{\mathbf{y}}_{1})}{1+(\tilde{\mathbf{y}}_{1}^{T}\tilde{\mathbf{y}}_{1})\left(\mathbf{g}_{1}+\mathbf{g}_{2}r_{12}\sqrt{\frac{\tilde{\mathbf{y}}_{2}^{T}\tilde{\mathbf{y}}_{2}}{\tilde{\mathbf{y}}_{1}^{T}\tilde{\mathbf{y}}_{1}}}\right)^{T}[\tilde{\mathbf{X}}_{nn}^{T}\tilde{\mathbf{X}}_{nn}+(\tilde{\mathbf{y}}_{2}^{T}\tilde{\mathbf{y}}_{2})(1-r_{12}^{2})\mathbf{g}_{2}\mathbf{g}_{2}^{T}]^{-1}\left(\mathbf{g}_{1}+\mathbf{g}_{2}r_{12}\sqrt{\frac{\tilde{\mathbf{y}}_{2}^{T}\tilde{\mathbf{y}}_{2}}{\tilde{\mathbf{y}}_{1}^{T}\tilde{\mathbf{y}}_{1}}}\right)} \tag{31}$$

shows that unspecific correlations *cannot* be avoided in the traditional method of calibration whenever $r_{12}^2>0$ because the correlated part of the other proteins will be added to the signal and subtracted from the noise. In order to produce a chemically specific calibration for albumin, the direct way of calibration must be employed, i.e., the spectral signal and the spectral noise must be estimated separately and the Wiener filter computed manually, with r_{12} set to zero. Incidentally, the fact that the glucose concentration in diabetic patients undergoes such violent and rapid swings is actually a *key* positive point from a calibration point of view because it allows the construction of chemically specific calibrations even when the shape of the glucose response spectrum is unknown. In many other applications, the issues in the future will be the following: Now that the math is spelled out and conscious decisions about the use of unspecific correlations can be made, will the various customers and regulatory agencies continue to be relatively forgiving for the use of unspecific correlations? Some intense discussions about the meaning of that phrase, ''specific in this application,'' can be expected in the future.

5.12 Which is Better, PLS or PCR?

There is no difference in quality between PLS and PCR. Any calibration is only as good as the SNR in the data, and that is what the algorithms use when they predict at their ''optimal'' ranks. Arguments are often construed that one is better than the other in terms of the number of factors necessary to build up a good **b** vector, but the relevance of that is very limited. Eliminating a PLS or PCR factor from the inversion is equivalent to defining the SNR_x in that spectral direction as zero. This can make perfect physical sense and can help the solution to become closer to the Wiener result, e.g., overfitting can be reduced by eliminating smaller eigenfactors that are known to represent nothing but electronic noise. The fact is, however, that elimination of factors does not help the hardware people. The Wiener **b** vector is what it is, and hardware will be needed to measure at all the pixels that span the **b** vector, with the SNR the **b** vector needs, whether or not somebody applies PLS or PCR. Setting the noise to zero in some mathematical subspace does not make the noise go away in reality. [Incidentally, all equations in this article also apply to rank-reduced inverses, e.g., when only the first PLS factors are used for inversion. In this case, the data in the complementary subspace (the unused factors) have to be thought of as effectively set to zero.]

5.13 Data Pretreatment

Data pretreatment methods are often claimed to improve the quality of calibration but in practice rarely do unless the measurement suffers from serious nonlinearity and/or nonstationarity problems, which many industrial process control applications do.[1,3] To repeat, the only thing that counts when solving Eq. (1) is the SNR, so the only pretreatments that have value are those that improve the SNR. When the data are linear and stationary, then there is no point in applying any more linear math to the spectra like first or second derivatives or other spectral filtering methods, because by definition they cannot improve upon what the optimum spectral filter aka Wiener filter will find in the data anyway. Like with PCR- or PLS-factor selection discussed above, there is limited use as a vehicle to insert *a priori* knowledge into the calibration and help the solution move closer to the Wiener result, but the result can never be better than what would have come from good calibration anyway. On the other hand, pretreatment methods that do more than just linear math on the spectra can potentially improve the SNR, e.g., the familiar spectral baseline correction methods (reduction of spectral noise in the larger eigenfactors).

5.14 Limitation Due to SNR_y

This limitation is best explained by using an example. Biomedical applications can be especially tough for many reasons. One important reason is that SNRs are typically limited by a lack of signal, because the concentrations in the human body hardly vary around their physiological averages to begin with. Established biomedical reference methods therefore work at SNRs of typically around 5, which is well above, but also not too far from, the cliff. If the goal is to develop a new method with, say, $\mathrm{SNR}=4$, compared to a reference method which supplies $\mathrm{SNR}_y=5$ to the new calibration, then the new method itself must measure with an $\mathrm{SNR}_x=6.8$ [Eq. (17)]. In other words, the closer the reference method pushes one to the cliff the harder it is to not fall down. This means that in many biomedical applications there is hardly any room left for losing correlation to the reference because of sample or sampling issues.

5.15 ''Classical'' Model

Any calibration method can be interpreted as an attempt to estimate the Wiener filter by looking at the mathematical details and analyzing what assumptions are implicitly made about the spectral signal and spectral noise which, when plugged into Eq. (13), give the particular method's prediction results. Consider the so-called classical model,

$$\tilde{\mathbf{x}}_{\mathrm{pred}}\equiv(\mathbf{g}\,\mathbf{K})\begin{pmatrix}\tilde{y}\\ \tilde{\mathbf{c}}\end{pmatrix}+\tilde{\mathbf{r}}, \tag{32}$$

where $\tilde{\mathbf{x}}_{\mathrm{pred}_{(kxl)}}$ is the column vector of new spectrum to be predicted (AU), $\mathbf{K}=[\mathbf{k}_1\,\mathbf{k}_2,\ldots,\mathbf{k}_R]_{(kxR)}$ is the matrix of interfering spectra or spectral effects (AU/mg/dL), $\tilde{\mathbf{c}}_{(Rxl)}$ is the vector of concentrations of the interferents (mg/dL), $\tilde{\mathbf{r}}_{(kxl)}$ is

the vector of residuals of the spectral fit (AU), and, as before, **g** (AU/mg/dL), the glucose response spectrum and $\tilde{y}$ (mg/dL) the sought-after (scalar) glucose concentration of $\tilde{\mathbf{x}}_{\text{pred}}$. For consistency in notation, Eq. (32) has been written in mean-centered form (where the mean spectrum is defined by the user) but this is not vital and could be dropped in the following discussion. The classical model is basically Beer's law in matrix notation. We assume here that *a priori* physical knowledge about the response spectra of the interfering components and other spectral effects, e.g., baseline variations, is available. Prediction of $\tilde{\mathbf{x}}_{\text{pred}}$ means that Eq. (32) is solved in a least-squares sense, yielding an estimate of the entire composition $[\tilde{y}\ \tilde{\mathbf{c}}^T]$ of the sample. The first question is, What is the equivalent **b** vector that the classical model uses to predict the glucose concentration $\tilde{y}$?

The LS solution of Eq. (32), which minimizes the SSE of the spectral fit, estimates the glucose concentration as

$$\tilde{y}_{\text{pred}}=\left\{(1,\ 0,...,0)\left[\begin{pmatrix}\mathbf{g}^T\\ \mathbf{K}^T\end{pmatrix}(\mathbf{g}\,\mathbf{K})\right]^{-1}\begin{pmatrix}\mathbf{g}^T\\ \mathbf{K}^T\end{pmatrix}\right\}\tilde{\mathbf{x}}_{\text{pred}}\equiv\hat{\mathbf{b}}_{\text{eq}}^T\tilde{\mathbf{x}}_{\text{pred}}, \tag{33}$$

where the vector (1, 0,...,0) is simply used to pick the glucose concentration out of the entire composition. Straightforward, yet tedious, algebra simplifies the above expression to

$$\hat{\mathbf{b}}_{\text{eq}}=\frac{[\mathbf{I}-\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T]\mathbf{g}}{\mathbf{g}^T[\mathbf{I}-\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T]\mathbf{g}}, \tag{34}$$

where matrix $\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$ is the projection matrix into the *R*-dimensional subspace spanned by the modeled interfering spectra. The next question is, What assumptions about SNR_x would a Wiener filter, Eq. (13), have to make in order to produce the **b**-vector result, Eq. (34)? Comparison of Eqs. (34) and (25) shows that the classical model is equivalent to a Wiener filter that (wrongly) assumes that

1. the SNR_x in subspace $\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$ is *zero* (no part of **g** in this subspace is used) regardless of the size of the amplitudes of the interfering spectra relative to the glucose signal; and
2. the SNR_x in orthogonal subspace $\mathbf{I}-\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$ is *infinitely good* regardless of the instrumental noise floor or, worse, any unmodeled interferents.

This is why the classical model has not performed well in demanding applications in the past and should generally not be used for concentration prediction. Whereas the result, Eq. (12), of the statistical model does converge against the Wiener filter for an increasing number of calibration samples, the result, Eq. (34), does *not* converge against the Wiener filter regardless of how much effort is put into estimating interfering spectra **K**. In fact, putting too much effort into defining **K** will invariably result into too large a number of modeled interferents and *degrade* prediction performance, because nothing of **g** is left to predict with. Besides, knowledge of the *individual* interfering spectra $\mathbf{k}_r$ $(r=1,2,...,R)$ is not needed anyway because only the projection matrix $\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$ is used for prediction. Recently, efforts have begun that are equivalent to moving the solution of the classical model closer to the Wiener filter,[26] but this approach has a long way to go and will be much harder in practice than the direct approach via the statistical model route.

Equation (34) was first presented to the chemometric community by Lorber[8] where it formed the basis for the net analyte signal concept. In later years, the NAS solution, Eq. (34), was erroneously claimed to also be the ''ideal'' result that the **b** vector of the statistical model converged against. However, the **b**-vector result of the statistical model, Eq. (1), is Eq. (12), which is completely different from Eq. (34). As far as prediction is concerned, classical modeling is basically identical to NAS calibration[9] and its various derivatives[10,11] with differences only in the definition of the projection matrix, $\mathbf{K}(\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$. This fact was recently pointed out by Kailey and Illing.[27]

5.16 Limit of Multivariate Detection

The spectral signal-to-noise ratio Eq. (14) can be written as

$$SNR_x=\frac{\sqrt{\dfrac{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}{m-1}}\cdot\|\mathbf{g}\|}{\sqrt{\dfrac{1}{\dfrac{\mathbf{g}^T}{\|\mathbf{g}\|}\left\{\dfrac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n}{m-1}\right\}^{-1}\dfrac{\mathbf{g}}{\|\mathbf{g}\|}}}}\left[\frac{(AU)_{RMS}}{(AU)_{RMS}}\right] \tag{14'}$$

$$SNR_x=\frac{\sqrt{\dfrac{\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}}{m-1}}}{\sqrt{\dfrac{1}{\mathbf{g}^T\left\{\dfrac{\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n}{m-1}\right\}^{-1}\mathbf{g}}}}\left[\frac{\left(\dfrac{mg}{dL}\right)_{RMS}}{\left(\dfrac{mg}{dL}\right)_{RMS}}\right] \tag{14''}$$

where the numerator is the rms signal and the denominator is the rms effective noise, in absorbance or in concentration units. The denominator of Eq. (14'') can be considered the limit of detection of the multivariate measurement. The rms prediction error ($\text{PRESS}^{1/2}$) in a scatter plot aproaches this value *if* spurious and unspecific correlations are zero and the slope is one and the reference noise is zero. The covariance matrix of the spectral noise transforms into the scalar effective noise in a peculiar way that is similar to a harmonic mean (''1 over inverse''), which is the mathematical reason why the effective noise is often much smaller than believed possible when looking at a measurement problem for the first time.

5.17 Economic Opportunities

The best thing about Eqs. (13) and (14) is that they mean net present value to companies because they point to a multitude of opportunities by which to reduce cost and time of development programs. Today's R&D efforts are characterized by a sequence of calibration experiments that are time consuming and expensive and, too often, inconclusive. The results in this article can be used to reduce the number of experiments needed to reach the goal. First, significant savings are possible whenever the pure component response spectrum (**g**) is known. All that is needed then is an estimate of the spectral noise and this may be possible under lab conditions, thereby avoiding the expense of collecting *in situ* calibration spectra.

The calibration **b** vector can be determined in a direct way by inserting the estimates for the spectral signal and spectral noise into Eq. (13), guaranteeing specificity and eliminating the danger of spurious correlations altogether. If the pure component response spectrum is not known, then many other opportunities still exist, especially when one good calibration experiment is performed as a starting point. The effect of additional noise sources on the SNR_x of an existing calibration can then be assessed quantitatively using Eq. (14). For example, imagine a single instrument is calibrated to a process and the task is to transfer this calibration to other instruments. One way to do this is to take a population of instruments and measure the instrument-to-instrument noise in the lab and on an average sample, then ''harden'' the existing calibration by adding the instrument-to-instrument noise to the spectral noise, and decide whether the hardened calibration still has enough SNR_x left. Other opportunities arise from intelligent optimization of the measurement hardware and process, always by assessing the effect on SNR_x in Eq. (14). This allows quantitative trade-offs between, e.g., the number of wavelength channels used and the final prediction correlation coefficient. It also improves communication between the hardware developers and the applications people by answering many of the questions from the hardware department without having to perform another calibration experiment, to an extent that a closed-loop feedback path can be established between hardware changes and system performance effects. It also avoids wasting time on ineffective issues like trying to improve the baseline stability of a glucose analyzer to μAU levels (the SNR_x in the spectral baseline direction is virtually zero because of varying amounts of interfering spectra from other blood components anyway). Instead, attention will be focused onto efforts that increase and protect the spectral directions with high SNR_x. In summary, there are a multitude of ways in which Eqs. (13) and (14) can bring significant savings to companies working in a number of different fields, and the amount of potential savings is great compared to the scale of the markets involved. A variety of very useful methods is described in a patent application by this author.

6 Example

The example chosen is the relatively simple case of the *in vitro* measurement of glucose in blood plasma in the mid-IR spectral range. A data set of 126 plasma samples from different, mostly diabetic patients was measured using an IFS-66 Fourier transform infrared (FTIR) spectrometer (Bruker, Karlsruhe, Germany) and an ATR micro-CIRCLE cell (Spectra Tech, Stamford, CT). The plasma samples were measured in a random sequence over a period of 8 days, including 6 measurement days. The reference concentrations of eight different analytes were determined, including glucose and total protein. Experimental details of this are given in Refs. 28 and 29. For our purposes here, no spectra are removed as outliers and the first 100 spectra collected are used to calibrate and the last 26 spectra collected are used as the prediction test set. The glucose calibration signal is $\sqrt{(\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R)/100} = 89.7$ mg/dL rms. Plasma absorbance spectra using water as the spectroscopic reference are shown in Figure 4. The measurement problem is slightly ill posed because the glucose spectrum is overlapped by other blood components. The standard deviation of the calibration spectra can be compared to the glucose signals shown in Figure 5, where trace C is the result of the author manipulating trace B to what he wished. The five absorbance bands between 1200 and 950 cm^{-1} are specific to glucose. Trace A in Figure 5 is the property-weighting spectrum (PWS) of the calibration spectra (200 mg/dL) $\times(\tilde{\mathbf{X}}^T\tilde{\mathbf{y}}_R)/(\tilde{\mathbf{y}}_R^T\tilde{\mathbf{y}}_R)$ and it has striking similarity to the pure glucose absorbance. However, some residual correlation to the protein bands in the 1500–1700 cm^{-1} range is also visible. The correlation coefficient between the glucose references and the total-protein references of the calibration samples was $r_{12} = 0.126$, which is very low and it is only

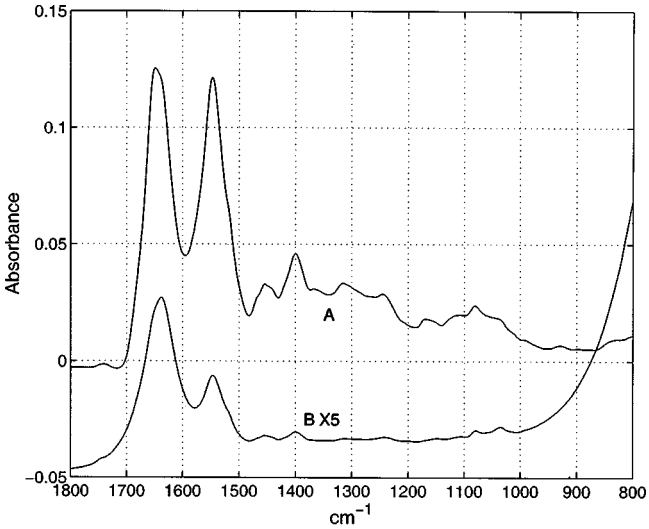

Fig. 4 ATR spectra of blood plasma in the mid-IR with water used as the spectroscopic reference: (A) average calibration spectrum and (B) standard deviation of the calibration spectra (enlarged and offset by −0.05 AU).

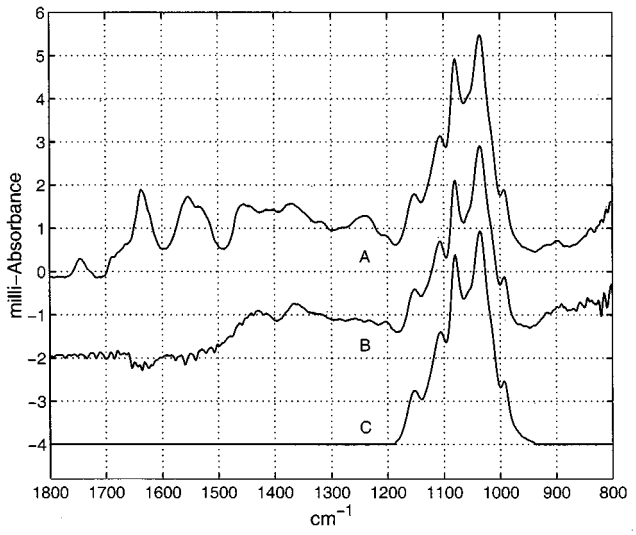

Fig. 5 (A) Property weighting spectrum of glucose; (B) spectrum of aqueous glucose solution (offset −2 mAU); and (C) user-manipulated spectrum of aqueous glucose solution (offset −4 mAU). All scaled to a concentration of 200 mg/dL.

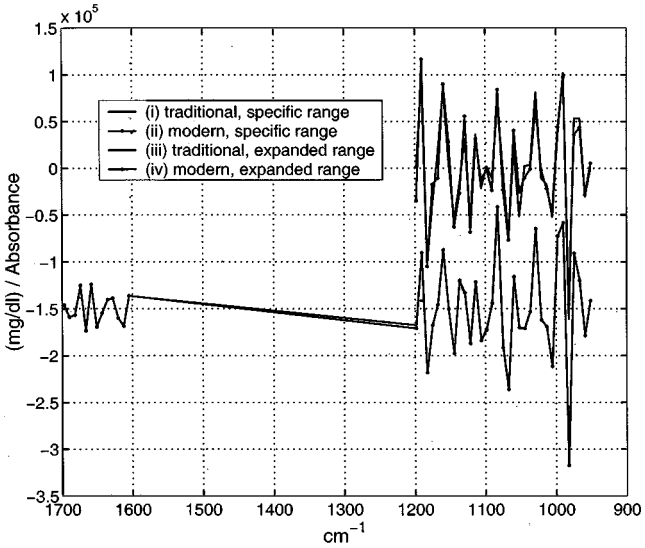

Fig. 6 **b** vectors of calibration scenarios (i)–(iv) described in the text. The results of the traditional and modern methods overlap almost perfectly. The **b** vectors in the expanded wavelength range are offset by −1.5e5 (mg/dL/AU).

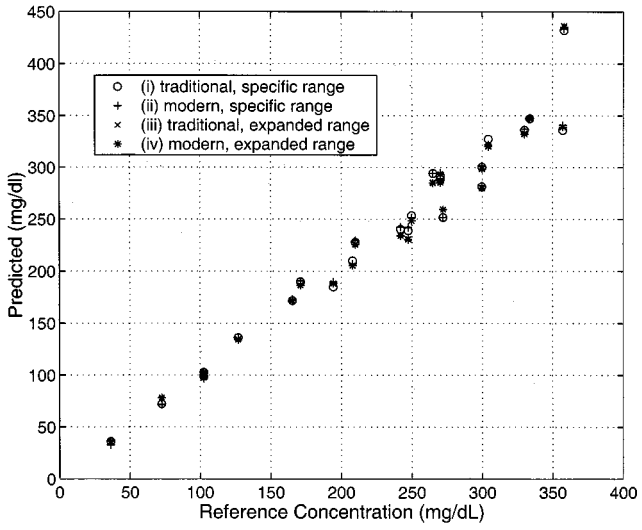

Fig. 7 Prediction scatter plots of scenarios (i)–(iv) described in the text.

because of the large amplitude of the protein absorbance bands that the PWS is visibly affected in the 1650 cm^{-1} region. In fact, the value of 0.126 is not even statistically significantly different from zero so that the correlation is very likely not a small unspecific correlation but, rather, a spurious effect of this calibration data set. Does this spurious correlation have any significant effect on the glucose prediction? This question is addressed below. Four different scenarios were used for calibration:

(i) the traditional way in the ''specific'' wavelength range=1198.2–951.4 cm^{-1} with 7.7 cm^{-1} intervals ($k=33$ channels);

(ii) the direct way in the specific wavelength range;

(iii) the traditional way in the ''expanded'' wavelength range=1198.2–951.4 *and* 1697.1–1604.5 cm^{-1} with 7.7 cm^{-1} intervals ($k=46$);

(iv) the direct way in the expanded wavelength range.

To repeat, the ''traditional'' way means that the data measured were plugged into Eq. (4) to generate the **b** vector, and the ''direct'' way means that the glucose spectral signal and spectral noise were estimated in a first step and then the estimates were plugged into Eq. (13) to generate the **b** vector. For the direct way, trace C in Figure 5 was used as the glucose response **g**; the reference concentrations $\tilde{\mathbf{y}}_R$ were used to estimate $\tilde{\mathbf{y}}$, and the difference $\tilde{\mathbf{X}}-\tilde{\mathbf{y}}\mathbf{g}^T$ was used to estimate the spectral noise. All matrix inversions used in this example are full rank, i.e., no PLS or PCR but plain least squares. The 1650 cm^{-1} region does not contain any glucose signal but was intentionally chosen to demonstrate the (insignificant) effect of the residual spurious correlation to the large protein bands. The **b**-vector results are shown in Figure 6 and the independent prediction results are shown in Figure 7. The results of the traditional and direct methods are virtually identical in this case of a well-designed calibration data set. The results for scenario (iii) demonstrate that the spurious correlation to the large protein bands around 1650 cm^{-1} seen in the PWS is not significant in the calibration, because correlation counts not covariance. Scenario (iv) in Figure 6 demonstrates that the nonzero wiggle of the **b** vector in the protein region is not caused by the spurious correlation there, but is caused by the true glucose signal when it is weighted down by the inverse matrix of the spectral noise.

Figure 8 is shown for completeness. Figure 8(a) shows the singular values of the 100 calibration spectra $\tilde{\mathbf{X}}$ in the expanded wavelength range and Figure 8(b) shows the correlation of the time eigenvectors to the glucose reference concentrations. The second and third factors clearly dominate the calibration. Figure 8(b) also shows that overfitting is no problem in this particular data set because the time correlations of eigenfactors higher than rank 10 are all virtually zero. Thus, using PCR or PLS to eliminate these eigenfactors would have been redundant here. Figure 8(c) shows the shape of the first few spectral eigenvectors and demonstrates the recovery of visualization possible in ill-posed measurement systems.

In Refs. 28 and 29 it was hypothesized that reference noise was the dominant contribution to the inaccuracy of the mid-IR glucose measurement because of the shape of the scatter plots, which showed large errors at the high concentration end and vice versa. The reference analyses were performed in triplicate at a certified clinical reference laboratory in the Diabetes Research Institute in Düsseldorf, Germany, which also ran ''gold standard'' controls at regular intervals. Based on the results of these controls and using a fairly detailed analysis,[28] the SNR_y of the calibration data set was estimated at a SNR_y of 9.8. We now quantify SNR_x by plugging the estimates of spectral signal and noise into Eq. (14), which gives a SNR_x of 12.5. The two results plugged into Eq. (17) show the total to be at a SNR of 7.7, which is excellent for a biomedical application. Figure 7 is the realization of a scatter plot with 26 points from a SNR of 7.7.

So, was it necessary to collect blood samples from diabetic patients to do this calibration? No. Since the response spec-

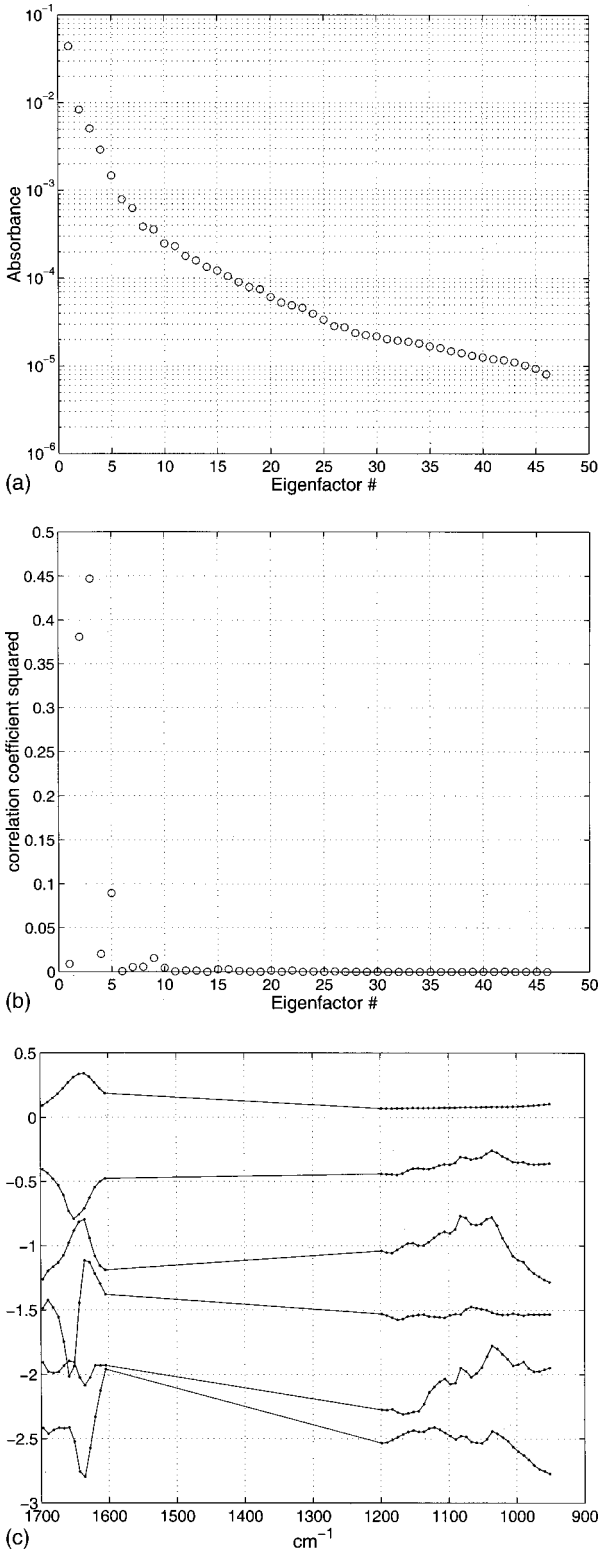

Fig. 8 Visualization of the ill-posed measurement problem. (a) Singular value decomposition $svd(\tilde{\mathbf{X}})/\sqrt{100}$ of the calibration spectra in the expanded wavelength range. (b) Correlation of the time eigenvectors to the glucose reference concentrations. (c) First six spectral eigenvectors, unitless, normalized to unity Euclidian length (No. 1 on top to No. 6 on the bottom; offset starts at zero and increases in steps of −0.5).

trum of glucose was known, the direct way of calibration could have been used and the spectral noise could have been measured from nondiabetic blood.

So, given that glucose is pretty easy to measure in the mid-IR, would not it be even easier to calibrate, say, albumin, which has much larger absorbance signals? As discussed in connection with Eq. (31) above, in the past the answer was ''no'' because of the strong unspecific correlations between the different proteins in the blood. Today, the answer is ''maybe.'' The fact that the absorbance values of albumin are large is good but has limited value in itself, because correlation counts not covariance. Sure, the albumin measurement is better posed than the glucose measurement and instrument noise is of less concern, however, whether or not the strongly overlapping spectra of the other proteins leave enough useful correlation aka SNR_x is up for grabs in a future study using the direct way of calibration. All results published so far based on the traditional method are corrupted by unspecific correlations.

If there is one important point to take away from Sec. 6 it is this one: **b** vectors are hard to interpret. Even though the mid-IR glucose measurement is simple (in the research lab) and only slightly ill posed, the **b** vectors in Figure 6 still do not look ''right.'' It is the spectral noise that makes them hard to interpret, though, even when the spectral signal has been very accurately identified. This author recommends moderation in trying to read physics from **b** vectors. For one thing, the human imagination does not work in >three-dimensional (3D) space. Also, instead of trying to visually solve the difficult inverse problem (Could this **b** vector be right for my analyte?), it is much easier to visualize the forward problem (OK, if this is my signal and this is my noise, then I guess this has to be my **b** vector.).

7 Summary

The so-called statistical calibration models are grounded on the physics of the pure component spectra. There are no fundamental differences between statistical and physical calibration models because both approaches are merely different attempts to realize the same basic idea, viz., to point the **b** vector into the direction with maximum spectral signal-to-noise ratio (SNR_x). This solution is the spectrometric Wiener filter and it is optimal in the mean-square prediction error sense. The rms pure component spectral signal $\sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m}\,\mathbf{g}\,(\mathrm{AU})$ and spectral noise $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m\,(\mathrm{AU}^2)$ are the two main physical building blocks that make up the spectrometric Wiener filter.

The closed-form solution, Eq. (12), of the statistical calibration model, Eq. (1), is given in terms of the pure component spectral signal, the spectral noise, the signal and noise of the reference method, and a scaling factor between the sample and reference concentrations. Equation (12) shows in detail how the traditional solution, Eq. (4), converges against the Wiener filter with an increase in the number of statistically independent calibration samples. Specifically, convergence requires that $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}}_n \rightarrow \mathbf{0}$, which means zero effect of reference noise; and $\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}} \rightarrow \mathbf{0}$ which means zero spurious correlations *and* zero unspecific correlations. Spurious correlations have been the biggest challenge and cost driver for many practical applications of multivariate calibration in the past. [For com-

pleteness, we should also state the obvious requirement that, in order to insure optimum performance in the *future*, the calibration signal $\sqrt{(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}})/m}\,\mathbf{g}$ and calibration noise $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m$ must also be good estimates of the true population statistics encountered in the future prediction spectra.]

The closed-form solution, Eq. (12), provides a wealth of practical benefits. First, it can be used to speed up the convergence against the Wiener filter. Second, it can be used to guarantee specificity. And third, it makes the calibration process fully transparent.

The ways by which to insert *a priori* knowledge are numerous and application specific, but the core statement is this: Different pieces of *a priori* physical knowledge about the spectra can be combined with any available measured data to estimate the pure component spectral signal and the spectral noise separately, and then compute the Wiener filter manually by plugging the results into Eq. (13). The effects of spurious correlations and reference noise are eliminated and the quality of the estimate of the Wiener filter is limited only by the quality of the initial estimates of the spectral signal and noise. In a fortunate case, where both $\mathbf{g}$ and $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m$ are known, collection of further calibration spectra is not necessary at all because the desired Wiener filter can be computed directly. In the more typical case, where spectral noise is not known, then calibration samples still have to be collected to estimate spectral noise, however, reference analyses are not necessary as soon as the shape of $\mathbf{g}$ is known. Trade-offs regarding practically important issues like calibration transfer or long-term stability can be made by adjusting the estimate of spectral noise. For example, a calibration can be made ''universal'' by including instrument-to-instrument and patient-to-patient noise.

The distinction between statistical and physical calibration models is artificial and should be a thing of the past. All calibration methods try to converge against the Wiener filter (or a ''slope-corrected'' version of it) and all use *statistical* estimates (from measured data) of the *physical* quantities that make up the Wiener filter. Practically speaking, it is often easier to approach the Wiener filter via the statistical model route rather than via the alternative classical model route because it is often easier to measure just the total rms spectral noise of a specific measurement application rather than to describe the noise in all its details. The Wiener filter requires knowledge of a single pure component response spectrum only (of the analyte of interest) whereas spectral noise can be determined as a total.

The signal-to-noise ratios of both the reference data (SNR_y) and the spectral data (SNR_x) were defined and the way in which they combine to form the total SNR of the application was given. Other statistics like the correlation coefficient, slope deficiency, scatter error, prediction error, etc., are highly nonlinear functions of SNR which, in turn, is really the only measure needed for assessing the quality of a measurement system and for decision making as to what needs to be done next in the development process.

The limited role of PLS and PCR was discussed. Also, the danger of spurious correlations in both the larger and the smaller eigenfactors was discussed in quantitative terms. The second most frequent reason for wild goose chase R&D efforts, the riding-the-cliff effect, was also discussed in quantitative terms.

The results in this paper can provide significant net present value to companies in various fields using multivariate calibrations, e.g., companies developing infrared spectrometric instruments and applications. Significant savings in cost and time for instrument calibration and calibration maintenance can be realized by reducing the number of expensive calibration experiments and by focusing hardware and process development efforts into areas that really count for system performance. The most important piece of physical information and the key to the most significant savings is knowledge of the shape of the pure component response spectrum of the analyte of interest. In addition, there is an opportunity for increases in revenue due to increased customer acceptance of calibration-based products.

Acknowledgments

The author thanks Dave Purdy, former President of Biocontrol Technology, Inc., Indiana, PA, for allowing him to grow in a challenging engineering environment. The author also thanks Augustyn Waczynski for being such a great engineering role model. The author also thanks Mike Heise of the Institute for Spectrochemistry and Applied Spectroscopy, Dortmund, Germany, for the teamwork during the PhD thesis. The Deutsche Forschungsgemeinschaft is thanked again for their financial support of the PhD work.

References

1. J. M. Schmitt and G. Kumar, ''Spectral distortions in near-infrared spectroscopy of turbid materials,'' *Appl. Spectrosc.* **50**, 1066–1073 (1996).
2. G. Kumar and J. M. Schmitt, ''Optimum probe geometry for near-infrared spectroscopy of biological tissue: Balancing light transmission and reflection,'' *Appl. Opt.* **36**, 2286–2293 (1997).
3. K. H. Norris and P. C. Williams, ''Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size,'' *Cereal Chem.* **61**, 158–165 (1984).
4. J. H. Williamson, ''Least-squares fitting of a straight line,'' *Can. J. Phys.* **46**, 1845–1847 (1968).
5. W. A. Fuller, *Measurement Error Models*, Wiley, New York (1987).
6. S. van Huffel and J. Vandewalle, *The Total Least Squares Problem. Computational Aspects and Analysis*, Society for Industrial and Applied Mathematics, Philadelphia (1991).
7. E. V. Thomas, ''Insights into multivariate calibration using errors-in-variables modeling,'' in *Recent Advances in Total Least Squares Techniques and Errors-in-variables Modeling*, S. van Huffel, Ed., pp. 359–370, Society for Industrial and Applied Mathematics, Philadelphia (1997).
8. A. Lorber, ''Error propagation and figures of merit for quantification by solving matrix equations,'' *Anal. Chem.* **58**, 1167–1172 (1986).
9. A. Lorber, K. Faaber, and B. R. Kowalski, ''Net analyte signal calculation in multivariate calibration,'' *Anal. Chem.* **69**, 1620–1626 (1997).
10. L. Xu and I. Schechter, ''A calibration method free of optimum factor number selection for automated multivariate analysis. Experimental and theoretical study,'' *Anal. Chem.* **69**, 3722–3730 (1997).
11. A. J. Berger, T. W. Koo, I. Itzkan, and M. S. Feld, ''An enhanced algorithm for linear multivariate calibration,'' *Anal. Chem.* **70**, 623–627 (1998).
12. D. H. Johnson, ''The application of spectral estimation methods to bearing estimation problems,'' *Proc. IEEE* **70**, 1018–1028 (1982).
13. J. C. Harsanyi and C. I. Chang, ''Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach,'' *IEEE Trans. Geosci. Remote Sens.* **32**, 779–784 (1994).
14. C. D. Brown, L. Vega-Montoto, and P. D. Wentzell, ''Derivative pre-

processing and optimal corrections for baseline drift in multivariate calibration,'' *Appl. Spectrosc.* **54**, 1055–1068 (2000).
15. H. M. Heise, ''Near-infrared spectrometry for *in-vivo* glucose sensing,'' Chap. 3 in *Biosensors in the Body. Continuous in Vivo Monitoring*, D. M. Fraser, Ed., pp. 79–116, Wiley, Chichester (1997).
16. G. H. Golub and C. F. van Loan, *Matrix Computations*, p. 3, The John Hopkins University Press, Baltimore (1983).
17. A. Papoulis, *Signal Analysis*, McGraw–Hill, Singapore (1984). Notice that the equations in this paper can all be written for the case of the ''general underlying'' distribution by replacing, e.g. $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{X}}_n)/m \rightarrow \mathrm{Cov}(\mathbf{x}_n)$ and $(\tilde{\mathbf{X}}_n^T\tilde{\mathbf{y}})/m \rightarrow \mathrm{Cov}(\mathbf{x}_n, y)$, etc.
18. T. Naes, ''Multivariate calibration when the error covariance matrix is structured,'' *Technometrics* **27**, 301–311 (1985).
19. R. Marbach and H. M. Heise, ''Calibration modeling by partial least-squares and principal component regression and its optimization using an improved leverage correction for prediction testing,'' *Chemom. Intell. Lab. Syst.* **9**, 45–63 (1990).
20. E. Stark, ''Near infrared spectroscopy past and future'' (quoting T. Hirschfeld), in *Near Infrared Spectroscopy: The Future Waves*, A. M. C. Davies and P. Williams, Eds., Proc. 7th Int'l Conf NIR Spectroscopy, Montreal, Canada, 6–11 August 1995, pp. 701–713, NIR, Chichester (1996).
21. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic, London (1979); see e.g., theorem 3.4.7.
22. D. E. Honigs, G. M. Hieftje, and T. Hirschfeld, ''A new method for obtaining individual component spectra from those of complex mixtures,'' *Appl. Spectrosc.* **38**, 317–322 (1984).
23. Test criterion established by R. A. Fisher, ''Frequency distribution of the values of the correlation coefficient in samples from an infinitely large population,'' *Biometrika* **10**, 507–521 (1915).
24. ASTM Standard E 1655-94, *Standard Practices for Infrared, Multivariate, Quantitative Analysis*, American Society for Testing and Materials, West Conshohocken, PA (1995).
25. T. W. Anderson, ''Asymptotic theory for principal component analysis,'' *Ann. Math. Stat.* **34**, 122–148 (1963).
26. D. M. Haaland and D. K. Melgaard, ''New prediction-augmented classical least-squares (PACLS) methods: Application to unmodeled interferents,'' *Appl. Spectrosc.* **54**, 1303–1312 (2000).
27. W. F. Kailey and L. Illing, ''Small target detection against vegetative backgrounds using hyperspectral imagery,'' 1997 Meeting of the IRIS Specialty Group on Passive Sensors, Vol. 1, pp. 423–429 (1997).
28. R. Marbach, *Messverfahren zur IR-spektroskopischen Blutglucosebestimmung*, PhD thesis, University of Dortmund, Germany, VDI, Düsseldorf (1993).
29. H. M. Heise, R. Marbach, Th. Koschinsky, and F. A. Gries, ''Multicomponent assay for blood substrates in human plasma by mid-infrared spectroscopy and its evaluation for clinical analysis,'' *Appl. Spectrosc.* **48**, 85–95 (1994).