JOURNAL OF CHEMOMETRICS
J. Chemometrics 2003; **17**: 646–652
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.832

Theory of net analyte signal vectors in inverse regression

Rasmus Bro[1]* and Charlotte M. Andersen[1,2]

[1]Food Technology, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark
[2]Danish Institute for Fisheries Research, Department of Seafood Research, DTU Building 221, DK-2800 Kgs Lyngby, Denmark

Received 19 September 2002; Revised 26 November 2003; Accepted 27 November 2003

The net analyte signal and the net analyte signal vector are useful measures in building and optimizing multivariate calibration models. In this paper a theory for their use in inverse regression is developed. The theory of net analyte signal was originally derived from classical least squares in spectral calibration where the responses of all pure analytes and interferents are assumed to be known. However, in chemometrics, inverse calibration models such as partial least squares regression are more abundant and several tools for calculating the net analyte signal in inverse regression models have been proposed. These methods yield different results and most do not provide results that are in accordance with the chosen calibration model. In this paper a thorough development of a *calibration-specific* net analyte signal vector is given. This definition turns out to be almost identical to the one recently suggested by Faber (*Anal. Chem. 1998;* 70: 5108–5110). A required correction of the net analyte signal in situations with negative predicted responses is also discussed. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: multivariate calibration; figures of merit

1. INTRODUCTION

In multivariate inverse calibration a regression model is sought predicting the response $\mathbf{y}$ of size $I \times 1$ from the multivariate measurements in $\mathbf{X}$ ($I \times J$). The regression vector $\mathbf{b}$ ($J \times 1$) is found to minimize the residuals $\mathbf{e}$ ($I \times 1$) in the equation

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

often subject to some additional constraints with respect to $\mathbf{X}$. For example, in principal component regression (PCR), $\mathbf{X}$ is constrained to be in the subspace of the first principal components, whereas in partial least squares regression (PLS) a slightly different criterion is used.

The net analyte signal vector is a convenient diagnostic that enables figures of merit (sensitivity, limit of detection, etc.) to be calculated in a manner similar to univariate regression. The net analyte signal vector was originally defined by Lorber [1] as the part of a measured signal that is orthogonal to the interferents. Although net analyte signals have mainly been discussed in relation to spectral data and especially in settings where Beer's law is assumed to be valid, the principle is general and both applicable and useful for any multivariate calibration model. In the following, though, the net analyte signal is discussed under the premise of spectral data.

The underlying idea of methods for calculating the net analyte signal is to separate the contributions in the calibration data matrix $\mathbf{X}$ into one originating solely from the analyte of interest (called $\mathbf{X}_k$ to reflect that the kth analyte is the analyte of interest) and another from other sources of variability such as interferents ($\mathbf{X}_{-k}$):

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} \tag{2}$$

$\mathbf{X}_k$ denotes the unique part of the analyte signal and $\mathbf{X}_{-k}$ is the matrix describing the signal orthogonal to that. Ideally, the unique analyte part lies in the null space of the space spanned by the spectra of the interferents. From a matrix spanning the space of the interferents ($\mathbf{X}_{-k}$), the net analyte signal vector of a sample i is calculated by

$$\mathbf{x}^*_{k,i} = [\mathbf{I} - (\mathbf{X}^+_{-k})\mathbf{X}_{-k}]\mathbf{x}_i \tag{3}$$

where $\mathbf{x}_i$ is the spectrum of the ith sample, $\mathbf{I}$ is a $J \times J$ identity matrix and $^+$ denotes a pseudoinverse. The net analyte signal is usually taken as the norm of $\mathbf{x}^*_{k,i}$ and can be used similarly to a univariate signal in univariate linear regression [2]. For e.g. mean-centered data a complication arises in calculating the net analyte signal. This will be illustrated when deriving a consistent algorithm for the net analyte signal vector.

The matrix $\mathbf{I} - (\mathbf{X}^+_{-k})\mathbf{X}_{-k}$ projects the calibration spectra onto the space orthogonal to that spanned by the spectra of

*Correspondence to: R. Bro, Food Technology, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.
E-mail: rb@kvl.dk

all analytes except the sought kth analyte. Thus, in order to find the net analyte signal vector of a certain analyte, it is necessary to find the projection matrix $\mathbf{I} - (\mathbf{X}_{-k}^{+})\mathbf{X}_{-k}$, which involves finding the matrix describing the interferent spectra, $\mathbf{X}_{-k}$. There are several ways to estimate this matrix.

Lorber *et al.* [3] suggested a method that uses PCR or PLS. First the calibration matrix $\mathbf{X}$ is rebuilt using A significant PCR or PLS components, yielding $\mathbf{X}_{\text{reb}}$. Then a rank annihilation step in the A-dimensional space is used for finding the part of the original matrix spanned by the interferents:

$$\mathbf{X}_{-k} = \mathbf{X}_{\text{reb}} - \alpha \mathbf{y}_k \mathbf{x}^{\mathrm{T}} \tag{4}$$

where $\mathbf{y}_k$ is the projection of the vector of responses $\mathbf{y}$ ($I \times 1$) onto the A-dimensional subspace and is given by $\mathbf{y}_k = \mathbf{X}_{\text{reb}}\mathbf{X}_{\text{reb}}^{+}\mathbf{y}$. The vector $\mathbf{x}$ is a linear combination of the rows of $\mathbf{X}$, which is chosen to include a contribution from the spectrum of the kth analyte. Any reasonable spectrum can be used for this purpose, though it is recommended to use a spectrum that contains maximal information on the analyte. The scalar α can be calculated as

$$\alpha = 1/\mathbf{x}^{\mathrm{T}}\mathbf{X}^{+}\mathbf{y}_k \tag{5}$$

and is simply included to account for any difference in magnitude associated with the non-related ways of determining $\mathbf{y}$ and $\mathbf{x}$.

Xu and Schechter [4] proposed another approach where $\mathbf{y}$ is used to define $\mathbf{X}_{-k}$. The calibration matrix $\mathbf{X}$ is scaled by dividing each spectral vector of matrix $\mathbf{X}$ by the corresponding y-value such that each spectral vector contains the same contribution of the analyte:

$$\mathbf{x}_{i,\text{sc}} = \frac{\mathbf{x}_i}{y_i} \tag{6}$$

In the next step the average of the scaled vectors is calculated and subtracted from all the scaled vectors. This gives a mean-centering pre-treatment of the scaled matrix removing the constant contribution of the analyte:

$$\mathbf{x}_{-k,i} = \mathbf{x}_{i,\text{sc}} - \mathbf{1}^{\mathrm{T}}\bar{x}_{\text{sc}} \tag{7}$$

where $\mathbf{1}$ is a J vector of ones. Combining $\mathbf{x}_{-k,i}$ for all samples provides an estimate of $\mathbf{X}_{-k}$. A similar approach is described by Goicoechea and Olivieri [5]. The mean calibration spectrum is obtained as in Equation (8) and the contribution of the analyte is subtracted from the data matrix $\mathbf{X}$ as shown in Equation (9):

$$\bar{\mathbf{x}} = \frac{1}{I}\sum_{i=1}^{I}\mathbf{x}_i \tag{8}$$

$$\mathbf{X}_{-k} = \mathbf{X} - \frac{\mathbf{y}\bar{\mathbf{x}}^{\mathrm{T}}}{\bar{y}} \tag{9}$$

where $\bar{y}$ denotes the mean calibration concentration of the analyte. Goicoechea and Olivieri [6] proposed to define $\mathbf{X}_{-k}$ as the projection of $\mathbf{X}$ orthogonal to $\mathbf{y}$ as illustrated in Equation (10).

$$\mathbf{X}_{-k} = [\mathbf{I} - \mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1}\mathbf{y}^{\mathrm{T}}]\mathbf{X} \tag{10}$$

Booksh and Kowalski [7] also described the net analyte signal vector and suggested a relation to the regression vector, but they did not provide an operational method for its calculation. Faber [8] put forward an idea which does not require the calculation of $\mathbf{X}_{-k}$. In this method the net analyte signal vector is calculated from the regression vector as

$$\mathbf{x}_{k,i}^{*} = \mathbf{b}(\mathbf{b}^{\mathrm{T}}\mathbf{b})^{-1}y_i \tag{11}$$

where y is the concentration of the analyte in the unknown sample (in practice, the prediction of the concentration is used). Originally the expected value of y_i was used, but in practice replaced with an estimate. This method was derived from the fact that the regression vector is the part of the analyte signal orthogonal to the signals of the interferents [9]. The rationale for this development was to circumvent the computational burden of some of the prior methods, and it was argued that the new method was an alternative that gave similar results to the older methods.

2. THEORY

Before deriving a net analyte signal vector, it is important to distinguish between two different ways of interpreting it. In his original publication, Lorber [1], introduced the net analyte signal as a practical tool to obtain figures of merit such as limit of detection. Such figures of merit are related to a specific model and in fact, one reason for obtaining figures of merit is to be able to compare different specific competing models.

Another usage of the net analyte signal has appeared based on the intuitive theoretical explanation behind it. In this alternative understanding, the net analyte signal is a unique property of a calibration *problem* which is independent of the estimated model. In this context, many different approaches can be reasonable for estimating the net analyte signal, but they are all interpreted as attempts to estimate the same underlying true net analyte signal.

In this newer interpretation of the net analyte signal the calibration problem is defined solely by the data set given. Underlying the assumption that only *one* net analyte signal can be defined is the fact that all calibration models aim at describing the same underlying 'true' regression vector. Mostly though, in practical inverse calibration, the calibration model depends critically not only on the data but also on the purpose of the model. Even though e.g. PLS is defined as the calibration model of choice and the data are given, the actual estimated (and also true) regression model can differ depending on the additional criteria to be fulfilled. These may for example be: to obtain the lowest possible prediction error variance in long-term predictions; to have optimal prediction errors in the short term; to have the best possible accuracy for high-concentration samples; etc. Clearly, such criteria have a real and meaningful effect on how many components can be extracted and hence on the estimated regression coefficients. Using the concept of *one* net analyte signal given a data set, such differences are not possible to quantify in terms of sensitivity, selectivity, limit of detection, etc.

In this paper we are not interested in the approach using the net analyte signal as a property of the data, even though it is fully valid and useful in certain areas. Rather, we focus

on the original understanding of the net analyte signal as a means for characterizing a specific analyzing system. Hence the net analyte signal we seek should be descriptive of the specifically chosen calibration model and not necessarily of the true net analyte signal (if one exists). If the calculated net analyte signal could estimate the true net analyte, it would be identical for any calibration model (ridge regression, principal component regression, etc.) and hence not enable comparison of different models from their different figures of merit.

In the above-defined context, most of the current approaches for calculating the net analyte signal vector suffer from the same problems. Unless noise-free data are used, the net analyte signal vector will not be proportional to the estimated regression vector. This proportionality should theoretically be present, because both the regression vector and the net analyte signal vector ideally represent the part of the space orthogonal to the interferents. Only then can they be representative of variations in the analyte only. This lack of equivalence between the estimated regression vector and the net analyte signal vector also has another practical aspect. When the vectors differ in shape, they will span different parts of the space, and hence any figure of merit calculated from the net analyte signal to represent the quality of the regression model may in fact be misleading and affected by variation that is not captured by the regression model. This situation is unsatisfactory both from a theoretical and a practical point of view when the net analyte signal approach is used to represent the regression model.

A further complication with some approaches is that the net analyte signal vector is to some extent independent of the calibration method. Hence, as also noted above, it can be impossible to compare different competing calibration models by their net analyte signal vector, as it will be the same regardless of the calibration method.

In this paper a simple calculation of the net analyte signal and its vector is derived. It has the advantages of not relying on any additional assumptions on the structure of the data and of providing a net analyte signal vector which will yield exactly the same predictions as the calibration model for which it is supposed to provide information. Furthermore, it handles mean centering automatically in a way similar to that suggested by Faber [10].

The matrices $\mathbf{X}_k$ and $\mathbf{X}_{-k}$ can be defined under the premises of a given (rank-reduced) calibration model in the following way. First of all, the null space of the model can be disregarded. For e.g. PCR the null space is given by the projection matrix $\mathbf{I}-\mathbf{PP}^+$, where $\mathbf{P}$ $(J \times F)$ is the F-dimensional loading matrix. This part of the space is not used for predicting and is regarded as noise in the model (used for diagnostic purposes). Instead of dividing the data into two different parts as is usually done in net analyte signal derivations, the model is divided into three contributions, i.e.

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} + \mathbf{X}_e \tag{12}$$

all orthogonal to each other. The matrix $\mathbf{X}_k$ is the part of the signal of the analyte which is orthogonal to the signal of the interferents and the noise, while $\mathbf{X}_{-k}$ is the part of the data in the subspace spanned by the interferents (including the part of the analyte signal within that subspace). The residual matrix $\mathbf{X}_e$ is e.g. defined as $\mathbf{X}(\mathbf{I}-\mathbf{PP}^+)$ for PCR but follows similarly for other calibration methods.

Within the space of the model given by $\mathbf{PP}^+$, the part given by the regression vector $\mathbf{b}$ is assumed to be in the same direction as the net analyte signal vector. Otherwise the predictions are influenced by other than analyte variation. Hence the null space of this direction must be equivalent to the space spanned by $\mathbf{X}_{-k}$. Therefore the matrix $\mathbf{X}_{-k}$ can be found from

$$\mathbf{X}_{-k} = \mathbf{XPP}^+(\mathbf{I} - \mathbf{bb}^+) \tag{13}$$

and hence

$$\mathbf{X}_k = \mathbf{XPP}^+\mathbf{bb}^+ = \mathbf{Xbb}^+ \tag{14}$$

because $\mathbf{b}$ is already defined as being in the column space of $\mathbf{P}$, hence $\mathbf{PP}^+\mathbf{b} = \mathbf{b}$. This means that the net analyte signal vector is simply found by projecting the data onto the space spanned by the regression vector. The main result in this paper is therefore that the net analyte signal vector of a certain sample, $\mathbf{x}_i$ is found as

$$\mathbf{x}_{k,i}^* = \mathbf{bb}^+\mathbf{x}_i = \mathbf{b}(\mathbf{b}^\mathrm{T}\mathbf{b})^{-1}\mathbf{b}^\mathrm{T}\mathbf{x}_i \tag{15}$$

where $\mathbf{b}$ is the regression vector from the calibration model. Note that estimated values are used throughout, because the interest is in assessing a specific (estimated) model, not a theoretical construct. The above definition of the net analyte signal circumvents the problems of the other methods and can be seen to be identical to the one proposed by Faber [8] when exchanging $\mathbf{b}^\mathrm{T}\mathbf{x}_i$ with y. Thus it retains the computational ease of his method. However, as opposed to earlier statements, it is argued that this is not merely another way to calculate the theoretically underlying net analyte signal vector. Rather, it is an explicit approach to estimate the net analyte signal vector of a regression model and which can be used to quantify the specific model's figures of merit.

The net analyte signal is usually defined as the norm of the net analyte signal vector. However, as pointed out by Faber [10], this is not valid for models that include centering. In fact, this definition is not even valid for non-centered models in situations where the response is predicted to be below zero (an example is given below). By properly deriving the calculation of the net analyte signal, the problem disappears. The prediction of y for a new sample $\mathbf{x}_i$ is defined as

$$\hat{y} = \mathbf{x}_i^\mathrm{T}\mathbf{b} = \mathbf{x}_{k,i}^{*\,\mathrm{T}}\mathbf{b} \tag{16}$$

because the net analyte signal is the part of $\mathbf{x}_i$ in the direction of $\mathbf{b}$. In the case of centered data, $\mathbf{x}_i$ corresponds to the centered data, and the prediction is of the centered response. This in itself poses no problems, but, given that the net analyte signal multiplied by the sensitivity $(1/\|\mathbf{b}\|)$ is supposed to equal the prediction of y, the definition of the net analyte signal $x_{k,i}^*$ follows as

$$\hat{y} = \mathbf{x}_{k,i}^{*\mathrm{T}}\mathbf{b} = \frac{x_{k,i}^*}{\|\mathbf{b}\|} \quad \Rightarrow \quad x_{k,i}^* = \frac{\mathbf{x}_{k,i}^{*\mathrm{T}}\mathbf{b}}{\|\mathbf{b}\|} \tag{17}$$

With this definition of the net analyte signal the magnitude remains the same as when calculated by the norm, but the sign is automatically set to be consistent with the obtained

Table I. Concentrations (μM) of amino acids in calibration samples

Sample	Tryptophan	Tyrosine	Phenylalanine
1	2.67	0.00	0
2	0.00	13.30	0
3	0.00	0.00	900
4	1.58	5.44	355
5	0.88	4.40	297

prediction of y. In practice, the correction above is equivalent to the one suggested by Faber [10], because the magnitude of the net analyte signal in Equation (17) remains unaltered whereas the sign changes with the sign of the prediction, as also implied in Faber's approach. It is emphasized, though, that it is necessary to use this approach even in situations where there is no centering involved.

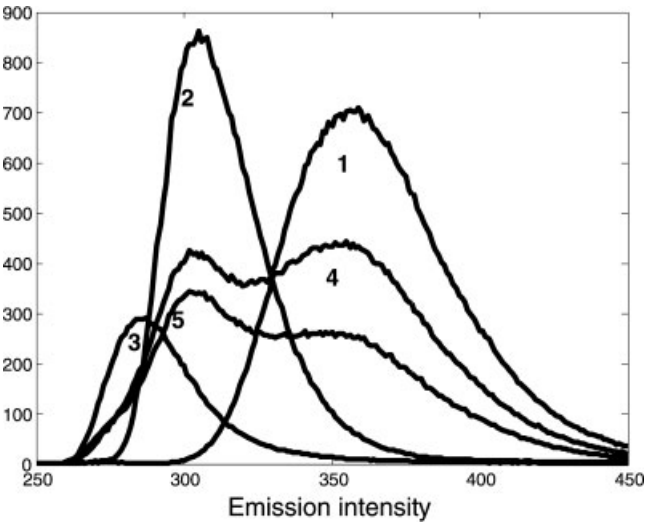

Figure 1. Emission spectra of the five samples.

3. RESULTS

The net analyte signal together with its vector developed here is very close to the one proposed earlier by Faber. A few illustrative examples will be provided for highlighting some of its properties compared with one of the earlier approaches. As an example of an earlier approach, the method of Lorber is chosen. This comparison is not meant to be exhaustive, as the earlier approaches have a different aim from the one presented here. They seek to estimate a net analyte signal vector assuming one exists. As explained, this is useful for other purposes not relevant here. Hence the comparison below mainly aims at showing that the two different kinds of approaches have different properties and that a regression model-specific net analyte signal vector is useful when assessing regression models.

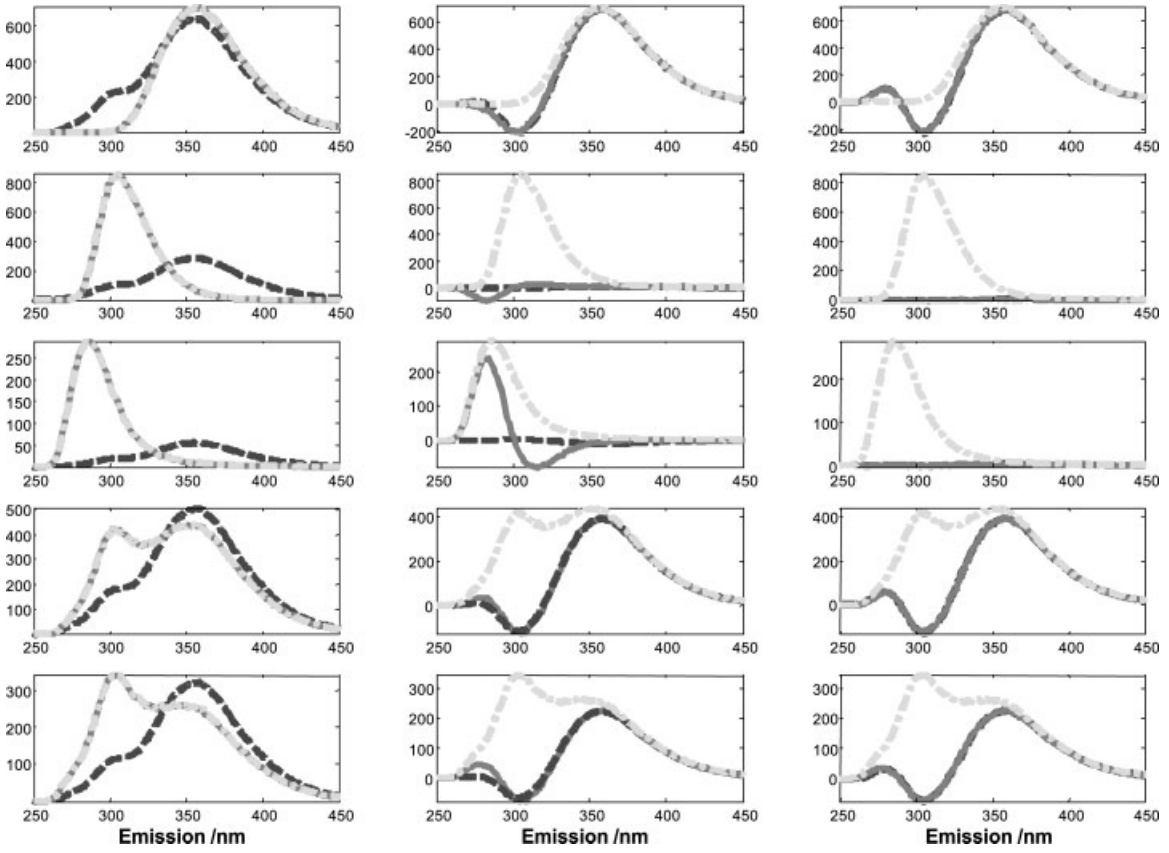

Figure 2. Results from a one-component (left), a two-component (middle) and a three-component (right) PLS model for tryptophan. Each row represents one sample and the chain line is the spectrum. The broken line is the new net analyte signal vector and the full line is the one calculated by Lorber's method. In the left and middle columns the full (Lorber) and chain (spectrum) lines are overlapping completely, whereas the full (Lorber) and broken (new) lines are overlapping completely in the right column.

Table II. Results from a two-component PLS model. Note the negative net analyte signal (NAS) of sample 3

Sample	True concentration	Predicted	NAS_{new}	NAS_{Lorber}
1	2.67	2.70	4.71	4.71
2	0.00	0.03	0.06	0.41
3	0.00	−0.04	−0.06	1.02
4	1.58	1.54	2.69	2.69
5	0.88	0.87	1.52	1.53

Two examples are given in the following. The first is a small and quite exaggerated illustration of some of the problems in earlier net analyte signal methods. The second provides a more typical example of a regression problem.

The first data set consists of five simple laboratory-made samples, each containing different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate-buffered water (Table I).

The samples were measured by fluorescence (excitation 269 nm, emission 250–450 nm, 1 nm intervals) on a PE LS50B spectrofluorometer with an excitation slit width of 2.5 nm, an emission slit width of 10 nm and a scan speed of 1500 nm s^{-1}. The data set is hence 5×201 (Figure 1).

A calibration model for tryptophan is made using PLS. Three components are expected to be optimal, but, in order to explore the net analyte signal vectors calculated by different methods, a one- and a two-component model are also evaluated. No centering is performed, as zero signal from an analyte is equivalent to zero concentration.

The results of using the new net analyte signal vector and the one proposed by Lorber (as implemented in the PLS_Toolbox version 2.1) are shown in Figure 2. For the three-component model there is virtually no difference between the two methods. For example, for samples 2 and 3, both methods suggest that the net analyte signal is zero,

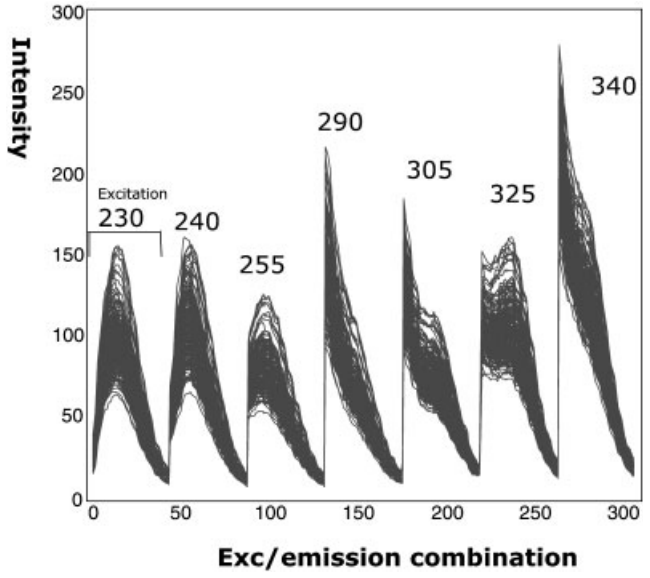

Figure 3. Fluorescence data from sugar example (five outliers removed). The seven excitation wavelengths (nm) used are indicated above the corresponding emission profiles.

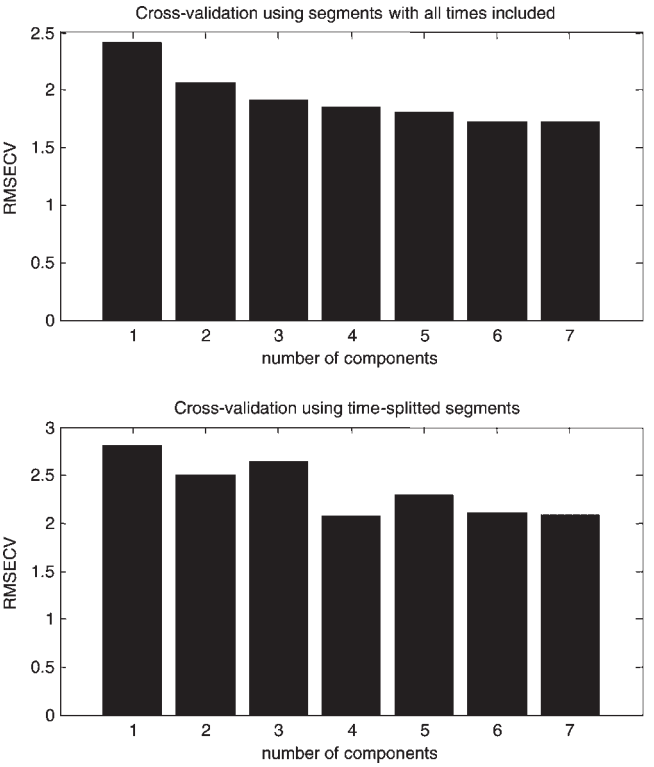

Figure 4. Cross-validation results using the same data but two different segmentations.

which is consistent with the zero concentration of tryptophan (Table I).

For the two-component model, however, the net analyte signals differ. For example, for samples 2 and 3 the new method suggests that the net analyte signal is still zero, while Lorber's method does not. Looking at the predictions of the two-component model (Table II), it is seen that indeed the predictions for samples 2 and 3 are close to zero. This is correctly reflected in the net analyte signal of the new method but not for Lorber's method.

For the one-component model, another interesting feature is clearly displayed. The net analyte signal from Lorber's method is completely overlapping the original spectrum, thus changing dramatically in shape from sample to sample. This follows from using Equation (3). This is counterintuitive, as the net analyte signal vector must be *one* specific direction independent of the actual sample. The same problem is not present with the suggested calculation of the net analyte signal vector.

The second example is from a larger set of samples of fluorescence measurements of sugar samples described by Bro [11]. The data can be found on www.models.kvl.dk (November 2003) and in this version the data consist of 268 samples of sugar measured at 230–340 nm excitation and 365–558.5 nm emission. Further details can be found in the original paper [11]. The data are rearranged to a two-way matrix of size 268×308 as shown in Figure 3.

The reference value to be predicted from these fluorescence data is the so-called color. A PLS model is built, but, before the number of components can be decided, the purpose of the model must be decided. The data are arranged in chronological order reflecting a 3 month period of

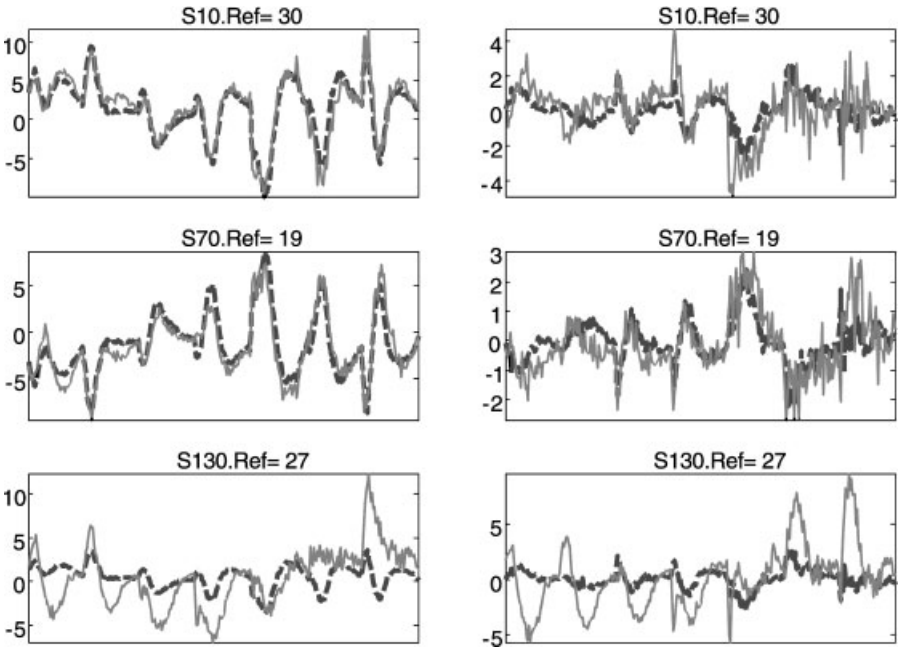

Figure 5. Net analyte signal vectors for a four-component (left) and a six-component (right) model. Vectors for three different samples are shown (Lorber method, full line; new approach, broken line). The reference variable color varies from 17 to 30 in the data. Note that the broken vectors remain the same shape in all samples (for each column); it is only the size and sign that change.

operation of the sugar factory from which the samples were obtained.

If the purpose of the model is merely internal validation to check whether there is any information in the fluorescence data on the color, then a sensible cross-validation procedure is to use a segmented validation where each segment contains samples from all periods. This is done using four segments, where the first left-out segment contains samples 1, 5, 9, etc.

If the purpose of the validation is to verify that the model can be expected to work in the future, then a useful (though not optimal) segmentation is to leave out segments in blocks. In this case the data are split into four segments, the first segment containing the first quarter of the samples, etc.

For these two types of validation the cross-validated results are given in Figure 4. Alternative cross-validation schemes using different numbers of segments were also tried, leading to similar results.

As can be seen, the choice of the number of components can vary depending on the purpose. To visualize some of the properties of the net analyte signal vector, both a four- and a six-component model are investigated. For these two models the net analyte signal vector was calculated using the approach suggested in this paper and the method of Lorber. For three randomly selected samples the results are shown in Figure 5.

The method of Lorber provides differently shaped net analyte signal vectors and some that deviate substantially from the current suggestion regardless of the number of components. The net analyte signal vector suggested in this paper remains the same shape and only the sign and magnitude change for a given regression model. For example, all leftmost vectors (four-component model) have the same shape for the suggested net analyte signal vector. Clearly, the differences in Lorber's method have implications for figures of merit and therefore for assessing a specific regression model's properties. As the current suggestion directly reflects the specific regression model, it is suggested that this makes it more adequate for assessing a specific regression model than e.g. Lorber's method. Furthermore, as stated earlier, an additional problem with methods such as Lorber's is that they cannot be used to assess differences in figures of merit between e.g. a PLS regression model and a ridge regression model, because they will provide the same figures of merit for both models even though this is obviously not reasonable.

4. CONCLUSION

A new consistent and general method for calculating net analyte signal vectors and net analyte signals in inverse calibration has been defined from a theoretical point of view. The theory and method are consistent with earlier literature for ideal cases but have several advantages in applications on real data.

REFERENCES

1. Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* 1986; **58**: 1167–1172.
2. Faber NM. Exact presentation of multivariate calibration model as univariate calibration graph. *Chemometrics Intell. Lab. Syst.* 2000; **50**: 107–114.
3. Lorber A, Faber NM, Kowalski BR. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* 1997; **69**: 1620–1626.

4. Xu LA, Schechter I. A calibration method free of optimum factor number selection for automated multivariate analysis. Experimental and theoretical study. *Anal. Chem.* 1997; **69**: 3722–3730.
5. Goicoechea HC, Olivieri AC. Enhanced synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Anal. Chem.* 1999; **71**: 4361–4368.
6. Goicoechea HC, Olivieri AC. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometrics Intell. Lab. Syst.* 2001; **56**: 73–81.
7. Booksh KS, Kowalski BR. Theory of analytical chemistry. *Anal. Chem.* 1994; **66**: 782A–791A.
8. Faber NM. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* 1998; **70**: 5108–5110.
9. Sanchez E, Kowalski BR. Tensorial calibration: I. First-order calibration. *J. Chemometrics* 1988; **2**: 247–263.
10. Faber NM. Mean centering and computation of scalar net analyte signal in multivariate calibration. *J. Chemometrics* 1998; **12**: 405–409.
11. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 133–147.