JOURNAL OF CHEMOMETRICS
J. Chemometrics 2005; **19**: 533–542
Published online 9 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.961

Applications of integrated sensing and processing in spectroscopic imaging and sensing

Joseph Medendorp and Robert A. Lodder*
Department of Pharmaceutical Sciences, College of Pharmacy, A123 ASTeCC Building 0286, Lexington, KY 40536, USA

Received 26 October 2005; Revised 27 December 2005; Accepted 20 January 2006

Integrated sensing and processing (ISP) encompasses the use of optical computing and adapted excitation signals to physically implement chemometric calculations in spectroscopic sensors for imaging. As data sets become larger and more complex with each emerging generation of hyperspectral imagers, the 'pixel-to-pupil' ratio increases at a rate faster than computing power can accommodate. In response to the need for faster and more efficient methods of processing, many analog solutions to the problem of high data dimensionality have emerged. The successful development of ISP has strong implications for military imaging, biosensing, spectroscopic imaging, and pharmaceutical process analytical technology (PAT). ISP developments in spectroscopy and PAT have emerged as alternatives to conventional Fourier transform infrared (FT-IR), near-infrared (NIR), IR, UV–visible, fluorescence, Raman, and acoustic-resonance spectrometry (ARS). Flourishing applications of ISP have demonstrated predictive ability equivalent to conventional approaches for sample differentiation and analyte quantification, in only a fraction of the time required for traditional spectrometric measurements. Copyright © 2006 John Wiley & Sons, Ltd.

KEYWORDS: hyperspectral integrated computational imaging (HICI), optics, computing, multiple linear regression (MLR)

1. INTRODUCTION

Integrated sensing and processing (ISP), a term coined by the Defense Advanced Research Projects Agency (DARPA), is an initiative calling for the implementation of new developments in chemometrics to the design of sensors and spectroscopic instrumentation [1]. Though not yet termed ISP, the idea was actually conceived in the 1970s and 1980s as a method to optically perform simple mathematical operations on spectroscopic data [2]. It has more recently been extended to more complex situations such as military imaging systems (in tanks and unmanned aerial vehicles) and the pharmaceutical industry for use in process analytical technology (PAT). PAT calls for the design and development of processes to guarantee a predefined quality of pharmaceutical materials at the end of the manufacturing process as warranted by risk analysis [3]. The goal of PAT is to encourage pharmaceutical companies to monitor each step of a manufacturing process, thus making ISP a perfect solution to the PAT initiative. ISP sensors are capable of measuring physical and chemical attributes in real-time, thus giving in-process and end-point knowledge of the quality and integrity of pharmaceutical materials [3]. Pharmaceutical PAT requires the development of rapid and accurate sensors so that existing full spectrum analytical instruments can be replaced with smaller, less expensive online alternatives. ISP is currently being studied as an alternative to traditional Fourier transform infrared (FT-IR), near-infrared (NIR), IR, UV–visible, fluorescence, Raman, and acoustic-resonance spectrometry (ARS).

Acoustic and optical spectroscopy coupled with multivariate mathematics is routinely used for analyte quantification and identification, such as in NIR imaging for the quantification of organics in the presence of interferents [4,5], scanning monochromator instruments and multichannel array spectrometers [6], and ARS for the classification of tablets and quantification of active pharmaceutical ingredient in semi-solids [7,8]. As the complexity and dimensionality of data sets increases with each emerging generation of hyperspectral imaging, the amount of computing power necessary for processing increases significantly. For example, each image collected with a digital micromirror array or an acoustic-resonance spectrometer may contain 500 000 discrete wavelengths or frequencies. When experimenters measure the correlation between spectral features and physical properties (e.g., concentration, temperature, and/or hydration), high-order relationships must be extracted. Often the relationship is determined with a method like

*Correspondence to: R. A. Lodder, Department of Pharmaceutical Sciences, College of Pharmacy, A123 ASTeCC Building 0286, Lexington, KY 40536, USA.
E-mail: Lodder@uky.edu
Contract/grant sponsor: National Institute of Health.
Contract/grant sponsor: National Institute on Alcohol Abuse and Alcoholism; contract/grant number: NIH NIAAA NO1 33003.

principal component analysis (PCA). Linear PCA is limited by the fact that the maximum number of principal components that can be extracted is $I-1$ where I is the number of spectra [9]. Because the number of wavelength or frequency observations often far exceeds the number of samples, these $I-1$ components can fail to identify the spectral features that correlate to the measured property of interest. Higher order algorithms such as support vector machines (SVM) and kernel PCA (KPCA) can be employed in such situations, such as multidimensional image and multivariate data analysis. SVMs have been used for analysis of mid-infrared spectra [10], nonlinear parametric models called multilayer perceptrons [11], NIR spectra affected by temperature-induced spectral variation [12], NIR spectra for acidity prediction in grapes [13], and neural networks [14]. KPCA has been used for optical character recognition and analysis of DNA [15], images [16], and NIR spectra [17].

While these approaches have proven successful, SVM and KPCA are computationally demanding. In ISP, rather than performing the calculations explicitly in a computer, some portion of the instrumentation (typically the detector) is designed to complement the spectral features of the incoming data set (e.g., frequencies and their respective amplitudes). The instrument itself extracts the distinguishing data structures without the need for further data processing. As new generations of spectroscopy and imaging arrive with higher dimensionality, detectors are built to share the computational burden, obviating the need for increases in post-collection computing power. Currently, there are many different research groups approaching this goal in a number of different ways.

1.1. ISP with optical signal processing

1.1.1. Theory

The most widely used ISP alternative to conventional optical spectroscopy involves the construction of optical interference filters. One of the first published methods, termed optical signal processing (OSP), is based on the four components illustrated in Figure 1: an element to disperse wavelength-dependent information into space-dependent information, a spatially variant optical transmission mask to filter incoming spectroscopic information, optical or analog elements to integrate the transmitted intensity, and a processor to perform simple algebraic computations [18]. The dispersion element directs wavelengths to different spatial locations on the transmission mask where the mathematical product is calculated between the spectroscopic information and the orthogonal weight function in the mask. The vector dot product is calculated as the sum or the integrated total signal by focusing the light from the transmission mask on the photodetector such that only one voltage, proportional to the concentration of the analyte, is recorded at the detector.

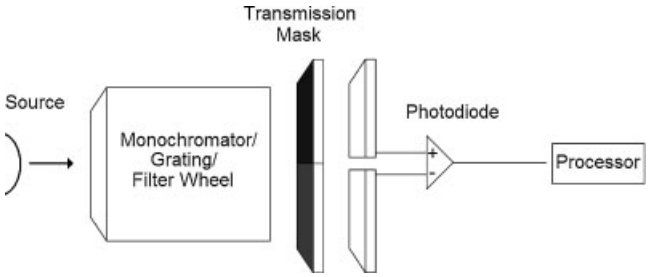

Figure 1. Optical signal processing (OSP) is based on the four components illustrated in this figure. The dispersion element (monochromator/grating/filter wheel) wavelength-dependent information is coupled to a spatially variant optical transmission mask to filter incoming spectroscopic information, optical or analog elements integrate the transmitted intensity, and the analog processor subtracts the signal from the negative lobe of the regression from the positive lobe signal.

To optically process an emission signal composed of several component spectra, a vector function is formulated so that it is orthogonal to the emission spectra of interfering species. The product of this vector and the interfering species effectively eliminates light emission from the interferents. The vector function is also formulated so that it is not orthogonal to the emission signal of the analyte of interest. In this manner the integral product of the vector and dispersed emission leaves only the signal from the analyte of interest, making the signal directly proportional to the analyte concentration, suggesting that the overall purpose of an OSP mask is to maximize the signal-to-background ratio. Clearly, this means that if there is significant overlap between the interferents and the analyte emission spectrum, or in the presence of many interferents, the throughput of the instrument will decrease significantly.

Mathematical construction of the optical filters can be illustrated in the following example. The total emission spectrum, free from interferences and directly related to the analyte property of interest is known as the net analyte signal (NAS) [19–22]. Free from instrument variation (noise, baseline shift) and chemical properties (temperature, viscosity etc.), the NAS can be represented by the wavelength-dependent emission spectrum, $\boldsymbol{a}_a$, and a total emission magnitude x_a. The interference components $\boldsymbol{a}_i$ and x_i are arranged in matrix $\mathbf{I}$. Gram–Schmidt orthonormalization of $\mathbf{I}$ results in a matrix of orthogonal vectors, $\mathbf{Q}$, related to the original matrix by the factorization matrix, $\mathbf{F}$.

$$\mathbf{I} = \mathbf{QF} \tag{1}$$

For an emission spectrum $\boldsymbol{a}$, a least-squares solution is used to determine the individual emission magnitudes, x_n, which are components of the vector $\boldsymbol{x}$, such that the 2-norm is a minimum:

$$\min\|\mathbf{I}\boldsymbol{x} - \boldsymbol{a}\|_2 \tag{2}$$

Premultiplication of Equation (2) gives:

$$\min\left\|(\mathbf{Q}^{\mathrm{T}}\mathbf{I})x - (\mathbf{Q}^{\mathrm{T}}\boldsymbol{a})\right\|_2 \tag{3}$$

where the superscript T indicates the matrix transpose. With Equation (1) and the fact that vector components of $\mathbf{Q}$ are orthonormal, the solution is:

$$\mathbf{F}\boldsymbol{x} = \mathbf{Q}^{\mathrm{T}}\boldsymbol{a} \tag{4}$$

$\mathbf{Q}^{\mathrm{T}}\boldsymbol{a}$ is a vector with elements equal to the dot products of $\boldsymbol{a}$ with the individual component vectors of $\mathbf{Q}^{\mathrm{T}}$. The dot

product of $\boldsymbol{q}_n$, the last vector of $\mathbf{Q}^{\mathrm{T}}$, with the emission spectrum, $\boldsymbol{a}$, is the optimal estimate, $\boldsymbol{x}_n$, which is used as the optical transmission filter.

Many recently published ISP applications stem directly from OSP research and the construction of optical interference filters. A discussion of the results appears later with the presentation of more recent OSP material.

1.2. ISP with optical regression

1.2.1. Theory

ISP with optical regression [23] is designed as an improvement over scanning spectrometers. In response to the assumption that in any data set there are many uninformative or useless variables, this method uses analog variable selection as opposed to post-collection processing, saving valuable time by collecting only the important variables. The overall purpose of optical regression is to improve S/N. Some authors contend that a digital regression, while powerful and flexible, does not necessarily give the optimal precision. In a typical scanning spectrometer, equal time is dedicated to the collection of each wavelength, thus valuable integration time is sometimes dedicated to the collection of useless data. The regression vector in optical regression serves as template for the amount of integration time spent at each wavelength. Thus, by varying the amount of time spent collecting light at each wavelength, the spectrum collected at the detector automatically represents the dot product of the transmission spectrum and the regression vector, saving valuable post-collection processing time. Because the transmission spectrum cannot be negative, two spectra are collected and differenced; one for the positive lobes of the regression vector, and one for the negative lobes. This difference is directly proportional to the concentration of the analyte of interest.

The integration times are determined by the inverse calibration model calculated by:

$$c = \boldsymbol{A}\boldsymbol{b} + \boldsymbol{e} \tag{5}$$

where $\boldsymbol{b}$ is the regression vector, c is a set of properties associated with an $I \times J$ matrix $\mathbf{A}$ composed of I independent samples and J observations, where $\boldsymbol{e}$ is a $J \times 1$ vector of model errors. The dot product is calculated between the regression vector and the unknown spectrum, $\boldsymbol{a}$:

$$\hat{c}_{\mathrm{un}} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = \sum_{j=1}^{J} a_j b_j \tag{6}$$

where $\hat{c}_{\mathrm{un}}$ is the estimation of the property of interest. If the regression vector is free from random errors, the errors in $\hat{c}_{\mathrm{un}}$ are the dot product of the errors, $\boldsymbol{e}$, in unknown spectrum $\boldsymbol{a}$ with the regression vector $\boldsymbol{b}$:

$$\boldsymbol{e}_{\hat{c}} = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{b} = \sum_{j=1}^{J} \boldsymbol{e}_j \boldsymbol{b}_j \tag{7}$$

The expected standard deviation of $\hat{c}_{\mathrm{un}}$ for normally distributed errors with mean of zero and standard deviation, σ, is given by:

$$\sigma_{\hat{c}} = \left(\sum_{j=1}^{J} \sigma_j^2 b_j^2\right)^{1/2} = \sigma\left(\sum_{j=1}^{J} b_j^2\right)^{1/2} = \sigma\|b\|_2 \tag{8}$$

In a spectrum with J wavelengths, there are J time units for the integration of the entire spectrum. The time dedicated for collection at each individual wavelength is determined by normalizing with the 1-norm, the sum of absolute values, on the regression vector to J:

$$b^t = b(J/\|b\|_1) \tag{9}$$

where $\boldsymbol{b}^t$ is the sum of $\boldsymbol{b}_{j+}^t$, the time dedicated to each of the wavelength with a positive regression value and $\boldsymbol{b}_{j-}^t$ the time dedicated to wavelengths with a negative regression value. The difference between the two measurements is equal to $J/\|b\|_1$ times the calibrated property.

$$\hat{c}_{\mathrm{un}} = \left(\sum_{j=1}^{J} \boldsymbol{b}_{j+}^t - \sum_{j=1}^{J} \boldsymbol{b}_{j-}^t\right)\|\boldsymbol{b}\|_1/J \tag{10}$$

If the measurement errors are equivalent to the errors in Equation (8), the standard deviation of $\hat{c}_{\mathrm{un}}$ is:

$$\sigma_{\hat{c}} = (\sigma_+^2 + \sigma_-^2)^{1/2}\|\mathbf{b}\|_1/J = 2^{1/2}\sigma\|\boldsymbol{b}\|_1/J \tag{11}$$

In digital regression, the detector collects K measurements per unit time over the J digitized channels, thus there are KJ total measurements. The regression vector must therefore be scaled such that the 1-norm equals KJ:

$$\mathbf{b}^t = \boldsymbol{b}(KJ\|\boldsymbol{b}\|_1 \tag{12}$$

where the J elements of $\boldsymbol{b}^t$ designate the number of signals integrated at each channel. The standard deviation of estimation here is expressed as:

$$\sigma_{\hat{c}} = (KJ)^{1/2}\sigma\|\boldsymbol{b}\|_1/J = \sigma K^{1/2}\|\boldsymbol{b}\|_1/J^{1/2} \tag{13}$$

The relative magnitudes of the standard deviations from analog (Equation 8) and digital (Equation 11) regression can now be compared. Analog optical regression predicts a lower standard deviation of estimation than digital regression by:

$$2^{1/2}\sigma\|\boldsymbol{b}\|_1/J < \sigma\|\boldsymbol{b}\|_2 \tag{14}$$

since by definition of multivariate analysis J is 2 or greater.

1.2.2. Results from optical regression

To demonstrate that optical regression outperforms digital regression in practice, Prakash et al. provide a short description of its application to the concentration prediction of four mixtures of rhodamine B, sulforhodamine, and rhodamine 590 in 20:80 methanol:water using fluorescence emission spectra. The regression vector was calculated by principal component regression (PCR), and visual inspection was used to select the 20 wavelengths with the highest predictive ability. The regression vector determined how much time was spent integrating the signal at each of the 20 wavelengths. With 20 wavelengths, digital regression allocated 30 milliseconds to integrate each channel, whereas optical regression integrated each channel according to its weight in the regression vector. Elimination or reduction of the contribution of less important wavelengths in optical

Table I. Optical regression demonstrated an average improvement over digital regression by 28.0% for the prediction of fluorescent dye mixtures

Analyte	Ex #	Conc. (ppm)	Mean bias (ppm)	σ_c digital	σ_c optical	Improvement (%)
Rhodamine B	1	5	−0.0074	0.0944	0.0700	25.8
	2	5	−0.0126	0.0828	0.0526	36.5
	3	0	0.0085	0.0880	0.0527	40.1
	4	5	0.0624	0.0815	0.0552	32.2
Rhodamine 590	1	1	−0.0248	0.0561	0.0527	6.1
	2	1	−0.0039	0.0438	0.0329	24.9
	3	1	0.0378	0.0570	0.0358	37.2
	4	0	−0.0502	0.0469	0.0364	22.4
Sulforhodamine	1	0	−0.0012	0.0620	0.0509	17.9
	2	10	0.0007	0.0617	0.0468	24.1
	3	10	−0.0526	0.0747	0.0458	38.7
	4	10	0.1089	0.0640	0.0450	29.7

regression effectively raised the signal-to-noise ratio. Of the 12 different mixtures, on average, optical regression predicted analyte concentration 28.0% more precisely than digital regression. Table I summarizes the results from optical regression.

1.3. Multivariate optical elements

1.3.1. Theory

The overall purpose of these optical interference filters, termed multivariate optical elements (MOE) [4–6,24–27], is to minimize the standard error of prediction (SEP), or the root-mean-squared error of calibration (RMSEC) of chemical species. MOE methods begin with the collection of spectra of target analytes. For example, if the objective is the NIR quantification of ethanol in water, spectra from different concentrations of ethanol in water are collected with a broadband source. Calculation of the principal components of the raw spectra is usually accomplished by a singular value decomposition of matrix **A** according to Equation (15):

$$\mathbf{A} = \mathbf{USL} \tag{15}$$

where **A** is the matrix of original spectra, the eigenvectors of $\mathbf{A}^T\mathbf{A}$ make up the columns of **L**, the eigenvectors of $\mathbf{AA}^T$ make up the columns of **U**, and **S** is a diagonal matrix of singular values [9].

As shown in Equation (16), a regression using **U** reveals which has the strongest correlation to a change in ethanol concentration c, where y is the y-intercept, $\boldsymbol{b}$ is a vector of regression coefficients, and $\boldsymbol{e}$, is the residual.

$$c = y + \boldsymbol{b}\mathbf{U} + \boldsymbol{e} \tag{16}$$

Equation (17) demonstrates how a leave-one-out cross-validation can be used in predicting the reliability of the concentrations of ethanol from NIR spectra, where $\boldsymbol{\sigma}^2$ is the variance, $\hat{c}_{un}$ is the prediction of the model for the ith pattern I in the training set, after it has been trained on the $I-1$ other patterns.

$$\sigma^2_{LOO} = \frac{1}{I}\sum_{i=1}^{i=I}(y_i - \hat{c}_{un})^2 \tag{17}$$

In the case of a minimal two-component system using ethanol and water, it is a simple matter to observe a linear change in ethanol concentration. With just two chemical constituents, only one principal component is needed to accurately describe the concentration changes that occur in mixtures. The loading corresponding to this principal component reflects the contribution of each wavelength to the overall classification. The concentration of unknown samples can be quantified by multiplying the loading and spectral intensity at the same wavelength, then summing over all wavelengths according to Equation (18) below:

$$y_i = \sum_{j=1}^{j=J} l_j \mathbf{A}_{ij} \tag{18}$$

where l_j is the loading at wavelength j, $\mathbf{A}_{ij}$ is the intensity of the raw spectrum of the ith mixture at wavelength j, y_i is the chemical analysis result, and j is an index over J wavelengths in the spectrum.

Figure 2 illustrates a generic version of the MOE selection process. The spectral variables demonstrate that the two peaks needed for sample differentiation describe the majority of the variation. The corresponding principal component loading identifies these regions and can be used as a template for the design of the optical filters. Because transmission spectra cannot be negative, negative loadings must be inverted as shown in MOE-2. Figures 3 and 4 illustrate the basic configurations of this instrument.

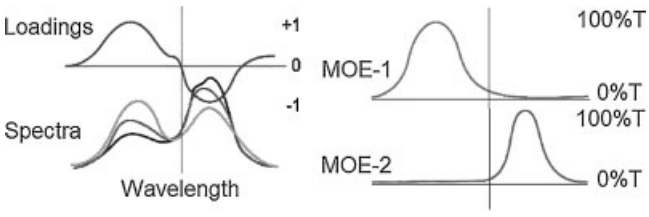

Figure 2. In this simple example, the spectral variables demonstrate that the two peaks needed for sample differentiation describe the majority of the variation in the spectra. The principal component loading corresponding to this change can be used as a template for the design of the optical filters. Molecular filters are designed so that their transmission spectra match the desired principal component-loading vector. Transmission spectra cannot be negative; therefore, negative loadings must be inverted separately as MOE-2.

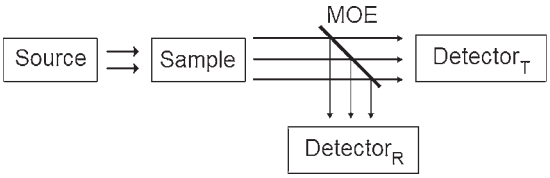

Figure 3. A simple T-format spectrometer with the multivariate optical element (MOE) as the beamsplitter. The transmitted signal passes through the MOE where it is collected at detector *T*, and the reflected signal bounces off the MOE for collection at detector *R*. The difference between transmission, *T*, and reflection, *R*, is directly proportional to the concentration of the analyte.

The MOEs in these configurations act as optical beamsplitters. The MOEs are designed so the transmittance to detector D_T at wavelength j is expressed by Equation (19), while the reflectance to detector D_R is expressed by Equation (20):

$$T_j = 0.5 + kl_j \tag{19}$$

$$R_j = 0.5 - kl_j \tag{20}$$

where k is a proportionality constant. These equations assume that the MOE does not absorb light or scatter stray light, thus the sum of T and R is unity at each wavelength. When light from all J wavelengths strikes the MOE, the total reflectance is given by Equation (21) and the total transmittance is given by Equation (22):

$$R_{t,i} = \sum_{j=1}^{j=J} (0.5 - kl_j)\mathbf{A}_{ij} = 0.5\sum_{j=1}^{j=J} \mathbf{A}_{ij} - k\sum_{j=1}^{j=J} l_j\mathbf{A}_{ij} \tag{21}$$

$$T_{t,i} = 0.5\sum_{j=1}^{j=J} \mathbf{A}_{ij} + k\sum_{j=1}^{j=J} l_j\mathbf{A}_{ij} \tag{22}$$

The difference between transmittance and reflectance is shown in Equation (23).

$$(T_{t,i} - R_{t,i}) = 2k\sum_{j=1}^{j=J} l_j\mathbf{A}_{ij} = 2ky_i \tag{23}$$

This equation demonstrates that the difference between transmittance and reflectance is directly proportional to analyte concentration.

The actual construction of the MOE has been approached in two different ways. The objective of both methods, however, is consistent; to design a MOE that has a transmission spectrum resembling the desired loading vector. In the first approach, a library of transmission spectra from organic compounds is screened to identify the optimum combination of organic materials whose transmission spectra will sum to the desired regression vector. The identified MOEs can be placed in a glass cuvette or in a rotating filter wheel directly before or after the analyte. These specific MOEs have been termed molecular optical filters. The second approach involves the deposition of alternating layers of two metal oxide films, Nb_2O_5 and SiO_2, on a 2.54 cm BK-7 glass substrate with reactive magnetron sputtering [28]. The interference effects of the layers are monitored as the layers are deposited to be sure the end product resembles the target transmission spectrum.

The utility of optical transmission filter spectrometers lies in the speed and accuracy at which they operate. Rather than using complex, bulky machinery with a dispersion element such as a grating or a stepping monochromator, one photodetector is capable of extracting the same information with no post-collection computation (the ISP advantage). Collection of all wavelengths simultaneously provides the multiplex advantage. The absence of a slit to achieve wavelength resolution provides the throughput advantage. An instrument simultaneously measuring a signal over a

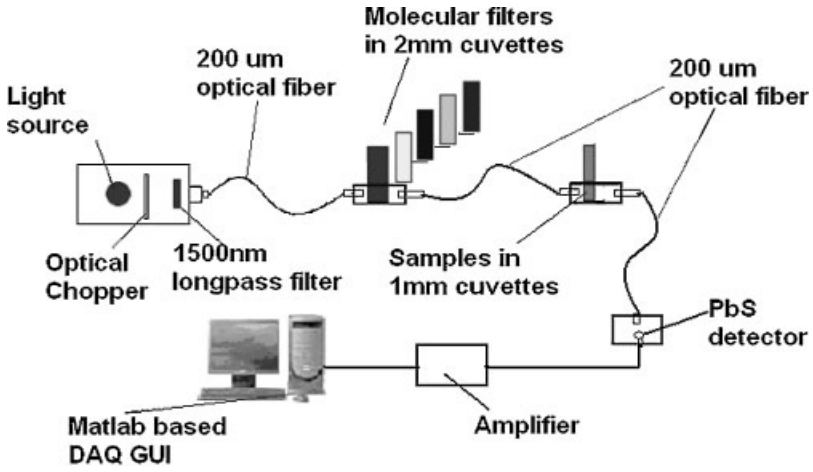

Figure 4. Diagram of a high-throughput molecular filter spectrometer. The molecular filters (MF), which are designed to resemble the principal component-loading vectors of interest, are placed in line with the sample. This implicitly calculates the dot product between the MFs and the sample transmission spectra, automatically weighting the most distinguishing variables for rapid and accurate analyte quantification.

range of wavelengths/frequencies obtains a $t^{1/2}$ advantage in the time t required to obtain a given signal-to-noise ratio compared to that which would be necessary using dispersive methods [29].

1.3.2. Select results from multivariate optical elements

Soyemi et al. describe an example of the use of a single MOE filter for the integrated sensing and prediction of analyte concentration. Transmission spectra from 400–650 nm were collected from mixtures of bismarck brown (BB) and crystal violet (CV), where BB was the analyte and CV was the random interferent. Digital principal component regression (PCR) using a four-component model gave a standard error of calibration (SEC) of 2.9% error relative to the mean of the calibration set, and a standard error of performance (SEP) of 4.3% relative to the mean of the calibration set. A plot of the root mean squared error of cross validation versus number of principal components indicated how many components were required for the best predictive ability. Ten MOEs, with layers ranging from 1 to 31, were constructed to resemble the regression vector as calculated from the PCR loadings. The best filter had 15 layers and gave SEC = 0.7% and SEP = 0.8% relative to the mean of the calibration set. (See Table II for the results of all the filters in comparison to the four-component digital regression model.) MOEs demonstrated an improvement of two orders of magnitude in both SEP and SEC over digital PCR for the prediction of analyte concentration [6]. The authors speculated that this is partly due to the inclusion of more model error in the PCR model, whereas the MOE avoids model error because it passes a smooth transmission waveform to the detector.

1.4. ISP in hyperspectral integrated computational imaging

1.4.1. Background

Hyperspectral integrated computational imaging is the imaging of a target analyte at a number of discrete wavelengths simultaneously [30]. As illustrated in Figure 5, HICI forms a data cube or hypercube. Data cube axes have different dimensions depending on the application. For example, the three dimensions in fluorescence excitation-emission matrix spectroscopy are excitation wavelength, emission wavelength, and sample number, which all serve as indices for the recorded property, spectral intensity [31]. However, for NIR imaging, spectral intensity is indexed by two spatial dimensions (length and width), excitation wavelength, and sample number [30,31]. Typical fluorometry and NIR spectroscopy are conducted by varying the excitation wavelengths and raster scanning across an analyte surface until an entire spectrum is collected at the photodetector. In HICI, data across all wavelengths are collected simultaneously with a digital camera such as a charge-coupled device (CCD) camera. In either case, the resulting image is often presented as a contour plot as illustrated in Figure 5. The collection of all data in one measurement speeds up data collection times and standardizes the results. For example, environmental conditions, temperature fluctuations, ambient light differences, and detector drift for a given set of scans may all be subject to change when collecting multiple spectra over time. On the other hand, these effects are negligible when collecting only one measurement. Thus, a main advantage of HICI over conventional imaging spectrometry is reduced data collection time.

Table II. Multivariate optical elements (MOE) with different numbers of layers constructed for the quantification of Bismarck brown and crystal violet dyes

No. of layers	SEC	SEP	Average RMS
29	0.00950	0.00792	0.00875
28	0.00924	0.00802	0.00865
19	0.00923	0.00796	0.00862
29	0.00841	0.00875	0.00858
20	0.00885	0.00822	0.00854
30	0.00919	0.00782	0.00853
21	0.00909	0.00784	0.00849
20	0.00879	0.00794	0.00838
25	0.00728	0.00846	0.00789
15	0.00699	0.00794	0.00748
PCR Model 4 components	0.029	0.043	0.045

Optical regression with the use of MOEs outperformed digital principal component regression by an order of magnitude for standard errors of calibration, performance, and the root mean squared errors. Statistics are reported relative to the mean of the calibration set, and as such, are unitless.

1.4.2. Theory

A number of chemometric techniques have been applied to the processing of the resulting data cubes from HICI images, including parallel factor analysis (PARAFAC) [31] and a nonparametric cluster analysis algorithm, the bootstrap error-adjusted single sample technique (BEAST) [32]. This review will concentrate on the BEAST, which can be implemented in simple hardware and is based on the premise that spectra from similar samples tend to cluster together in the same region of hyperspace. The BEAST draws probability-density contour plots around analyte populations based on asymmetric standard deviations (SDs) [32]. For a more detailed description of the BEAST, an in-depth discussion is provided by Hamilton et al. In the BEAST, a population $\mathbf{P}$ is created as an $I \times J$ matrix in hyperspace H whose rows are the individual samples and the columns are the wavelengths. $\mathbf{P}^*$ is a discrete realization of $\mathbf{P}$ based on a calibration set A of the same dimensions as $\mathbf{P}^*$. This realization is chosen one time from $\mathbf{P}$ to approximate all possible sample variations present in $\mathbf{P}$. $\mathbf{P}^*$ has parameters M_c and $\overline{\mathbf{P}}$, where $\overline{\mathbf{P}} = E(\mathbf{P})$ and M_c is the Monte Carlo approximation to the bootstrap distribution. The expectation value, $E(\mathbf{P})$, is the center of $\mathbf{P}$, and $\overline{P}$ is a row vector with the same number of rows as there are columns in vector $\mathbf{P}$. New test spectra A_T are projected into H containing M_c, and rows of M_c are mapped onto a vector connecting $\overline{\mathbf{P}}$ and A_T. $\overline{\mathbf{P}}$ and A_T have the same dimensions. The integral over H is calculated from the center of $\mathbf{P}$ in all directions. A skew-adjusted SD is based on the comparison of the expectation value $\overline{\mathbf{P}} = E(\mathbf{P})$ and $\overline{\mathbf{P}} = \mathrm{med}(A)$, the median of A in hyperspace is projected on the hyperline connecting $\overline{\mathbf{P}}$ and A_T. The result is an asymmetric SD that provides two

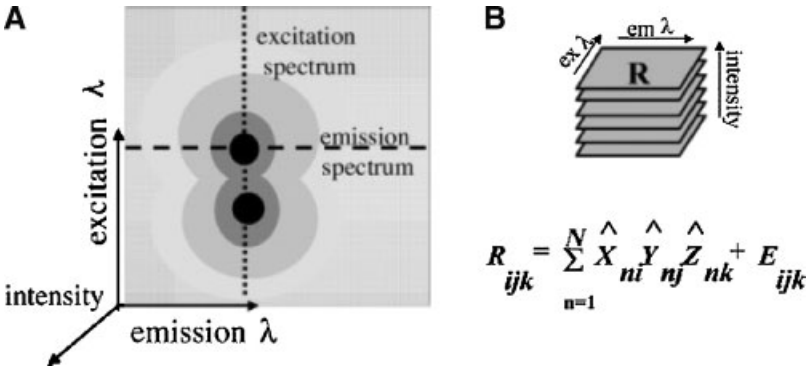

Figure 5. (a) A contour plot of an excitation-emission fluorescence spectrum. Rows represent emission spectra at a particular excitation wavelength, and columns represent excitation spectra at a particular emission wavelength. (b) A three-way data cube can be constructed by stacking correlated EEM spectra. The cube can be mathematically modeled by a set of trilinear components, where R_{ijk} is the fluorescence intensity of sample k at excitation wavelength i and emission wavelength j, N is the number of unique spectral profiles found in the data cube, the columns of X, Y, and Z are the estimations of the pure excitation, emission and concentration profiles, respectively and $\mathbf{E}_{ijk}$ is the error matrix.

measures of the SD along the hyperline connecting $\overline{\mathbf{P}}$ and A_T. Skew adjusted SDs can be used to calculate mean distances between spectra of different samples.

1.4.3. Select results from hyperspectral integrated computational imaging

Groups of aspirin tablets exposed to different levels of moisture have been used to test the assumption that the BEAST approach offered an advantage over conventional NIR spectroscopy [30]. Tablet groups were exposed to water vapor or a pH 9.0 ammonium hydroxide solution. Water uptake was determined by gravimetric analysis and NIR spectrometry. To verify that NIR could be used to differentiate between different tablets, calibration lines were constructed for both water uptake and for the decomposition of aspirin into salicylic acid. The standard error of estimate (SEE) for the prediction of water content was 0.05% and the SEP was 0.06%. The SEE and the SEP for the prediction of salicylic acid were each 0.06%. NIR images of 1286 control tablets were collected simultaneously with a CCD camera, and the BEST algorithm was used to calculate multi-dimensional standard deviations. Figure 6 is a contour plot with lines drawn in BEAST standard deviations. It is apparent from this plot that NIR imaging can readily indicate water exposure time. In this example, NIR imaging resulted in a 30 000-fold increase in speed over HPLC and a 1000-fold increase in speed over conventional NIR spectrometry, suggesting a high degree of utility in PAT, as thousands of tablets can be scanned simultaneously and in real-time as they move down a manufacturing line.

1.5. Acoustic-resonance spectrometry with an ISP excitation signal

1.5.1. Theory

Typical acoustic-resonance spectrometric measurements involve very large amounts of data. For example, 15 seconds of data collection at a sample rate of 44.1 kHz results in 661 500 data points per spectrum. Normally, there is a large disparity between the number of sample spectra and the number of frequencies at which data are collected on the samples [33]. Due to instrumental (noise, baseline shift etc.) and chemical inconsistencies (temperature, viscosity etc.), spectral features of interest may be severely overlapped by interferences. Additionally, not all frequencies respond linearly or exclusively to the desired change in analyte concentration, therefore, linear principal component regression may be inadequate to describe the relationship between

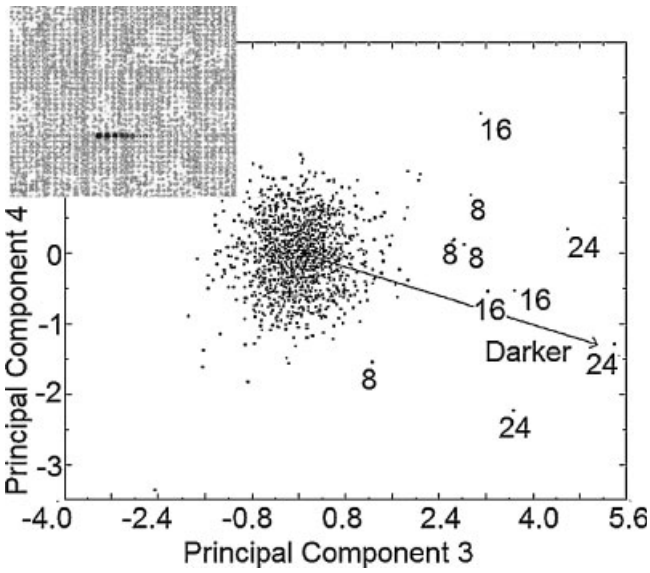

Figure 6. Bootstrap error-adjusted single sample technique (BEAST) contour plot of 1286 standard aspirin tablets scanned through the blister packaging with a NIR camera. Distances calculated in BEAST standard deviations from the center of the population distribution indicate the moisture absorption of individual tablets.

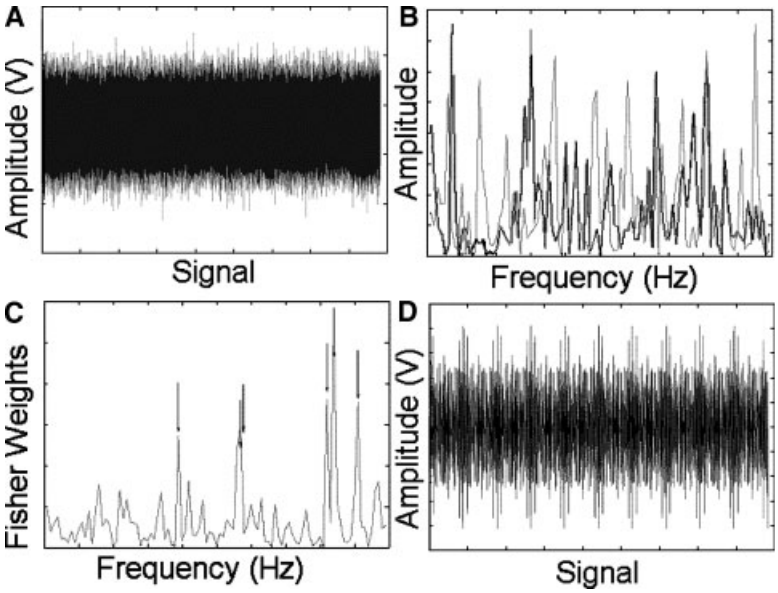

Figure 7. (A) Voltage signal collected for each analyte [sample rate of 44.1 kHz, duration 15 seconds, thus 661 500 data points]; (B) Fourier transform calculated to put data into frequency domain, allows for identification of distinguishing frequencies, thin line, analyte 1; thick line, analyte 2; (C) calculation of Fisher weights identifies most distinguishing frequencies; and (D) those frequencies are used to construct the sound file, (for simplicity of visualization, the sum of only six frequencies is used in this example), using a sample rate of 44.1 kHz for 15-second duration.

AR spectra and physical properties. For PAT applications the data burden mandates that an ISP method be developed to reduce the dimensionality, retaining only those frequencies directly proportional to the concentration of the physical property of interest.

Figure 7 illustrates the ISP encoding and processing scheme. Full spectra are collected in the time domain, and transformed to the frequency domain with an FFT. Fisher weights are calculated for each discrete frequency according to Equation (24) [7,8]. Frequencies with the top weights are selected, and an inverse Fourier transform is performed to put the frequencies back into the time domain. This file, which serves as the new excitation signal, can be saved as an MP3 file and encoded on a CD/MP3 player.

$$\mathbf{F} = \frac{\left|\overline{\mathbf{P}}_1 - \overline{\mathbf{P}}_2\right|}{\left(\frac{1}{I-1}\sum_{i=1}^{i=I}(\mathbf{P}_{1i} - \overline{\mathbf{P}}_1)\right) - \left(\frac{1}{I-1}\sum_{i=1}^{i=I}(\mathbf{P}_{2i} - \overline{\mathbf{P}}_2)\right)} \tag{24}$$

Because the excitation signal was created using only those frequencies *directly proportional* to the concentration of the physical property of interest, once the signal passes through the sample the signal intensity is directly proportional to the analyte concentration. No FFT or spectral sweeping is required to generate the result. In fact, a simple rectifier circuit and voltmeter will suffice [8].

In generating the Fisher weights, data populations $\mathbf{P}_1$ and $\mathbf{P}_2$ contain $I \times J$ points where I rows contain the samples and J columns contain the frequencies. The algorithm uses an $J \times 1$ classification matrix, $\mathbf{c}$, to identify the number of unique classes, $\mathbf{c}_u$, in a given data set. The number of possible combinations of classes, and thus the number of rows in the Fisher weights matrix $\mathbf{F}$, I_f, is calculated by Equation (25).

$$I_f = \frac{1}{2}\mathbf{c}_u(\mathbf{c}_u - 1) \tag{25}$$

Each row in $\mathbf{F}$ is composed of short sound segments or 'bites' that are designed for the differentiation of specific analytes. When trying to differentiate between groups of similar samples, the most distinguishing frequencies for one analyte are usually the most distinguishing frequencies for all analytes in that group. It is this property that allows the use of only one row in $\mathbf{F}$ for the final ISP sound segment. The row is selected based upon some other method of multivariate analysis, such as partial least squares, PCR, or the BEAST nonparametric clustering algorithm. Whichever two analytes exhibit the largest separation according to these methods indicates which of the rows in $\mathbf{F}$ are used as the excitation source.

Despite the potential introduction of a few uninformative variables for each individual calibration by using only one row in $\mathbf{F}$, the overall reduction in insignificant frequencies boosts the S/N much higher than seen in conventional ARS. In addition, the ISP approach offers a large advantage in terms of speed, which outweighs the small loss in performance due to the possible inclusion of useless variables. Data sets collected with the new excitation source are presently still collected in full, however, the computational burden is greatly reduced by filtering the incoming data according to the pre-selection of the distinguishing frequencies. Future AR spectrometers will include analog

Table III. Bootstrap error-adjusted single sample technique (BEAST) standard deviations for tablet clusters as scanned by acoustic resonance spectrometry (ARS)

	Aspirin	Ibuprofen	APAP	Vitamin B12	Vitamin C
NonISP					
Aspirin	X	19.17	5.32	1.84	6.13
Ibuprofen		X	12.30	29.13	30.68
APAP			X	10.09	16.10
Vitamin B12				X	4.47
Vitamin C					X
ISP					
Aspirin	X	3.41	4.01	5.09	15.03
Ibuprofen		X	6.98	14.07	21.17
APAP			X	10.26	16.13
Vitamin B12				X	12.94
Vitamin C					X

The nonintegrated sensing and processing (ISP) approach resulted in an average intertablet standard deviation of 13.52 and the ISP approach had an average intertablet standard deviation 10.8. Aspirin and vitamin B12 failed to separate in the nonISP approach, whereas all tablets separated in the ISP approach. Therefore, ISP offers both an advantage in separation as well as faster processing times.

frequency selection elements and an analog integrator so that fewer data points can be collected.

1.5.2. Results from ISP acoustic-resonance spectrometry

ISP ARS has been applied to the differentiation of similarly sized and shaped tablets [7,8,33]. The success metric for the differentiation of tablets was intercluster distance in BEST standard deviations. Five groups of tablets, aspirin, ibuprofen, acetaminophen, vitamin B12, and vitamin C were selected for analysis. For initial data collection, tablets were scanned in random order for 15 seconds at a sample rate of 8 kHz. BEAST SDs were calculated, and the excitation signal was selected from the tablets with the largest intergroup separation. All tablets were rescanned with the new excitation source for the ISP approach. The average multidimensional SD between tablet groups for the nonISP approach was 13.52, and for the ISP approach was 10.8. Table III lists all results. While the average SD was higher for the nonISP approach, tablet populations separated by three or more SDs are considered separable. In addition, not all tablet groups separated in the nonISP approach. This suggests that the ISP offers an advantage in reliability of separation as well as in data reduction and simplicity of detection, presumably by the increase in S/N due to the elimination of uninformative variables.

2. CONCLUSION

ISP is a viable alternative to conventional spectroscopy and imaging, offering substantially faster analysis times while maintaining equivalent performance with its digital counterparts. Without ISP, it is necessary for analytical techniques to make a trade-off between longer data acquisition (scan) times or lower predictive ability. With the emergence of ISP, however, experimenters can reduce the data collection times and data dimensionality without forfeiting high-predictive performance. ISP has shown promising results in spectroscopy and imaging, and has demonstrated a high degree of utility for military applications and pharmaceutical PAT.

REFERENCES

1. DARPA Defense Sciences Office. Integrated sensing and processing. Retrieved 21 May 2005. http://www.darpa.mil/dso/thrust/math/isp.htm/.
2. Pantelic D. Optical computation of determinants. *Optics Commun*. 1987; **64**: 421–424.
3. Process Analytical Technology (PAT) Initiative.U.S. Food and Drug Administration. Retrieved 1 April 2005 http://www.fda.gov/cder/OPS/PAT.htm/.
4. Myrick M, Soyemi O, Haibach F, Zhang L, Greer A, Li H, Priore R, Schiza M, Farr J. Application of multivariate optical computing to near-infrared imaging. *Proc. SPIE* 2002; **4577**: 148–157.
5. Myrick M, Soyemi O, Schiza M, Farr J, Haibach F, Greer A, Li H, Priore R. Application of multivariate optical computing to simple near-infrared point measurements. *Proc. SPIE* 2002; **4574**: 208–215.
6. Haibach F, Myrick M. Precision in multivariate optical computing. *Appl. Optics* 2004; **43**(10): 2130–2140.
7. Medendorp J, Lodder RA. *Integrated Sensing and Processing and a Novel Acoustic-resonance Spectrometer*. American Association of Pharmaceutical Sciences: Baltimore, MD, 2004.
8. Medendorp J, Lodder RA. Integrated sensing and processing with a tailored excitation signal for acoustic-resonance spectrometry. The Pittsburgh Conference, Orlando, FL: 2005.
9. Jolliffe IT. *Principal Component Analysis*. Springer: New York, 2002.
10. Belousov A, Verzakov S, von Frese J. Applicational aspects of support vector machines. *J. Chemometrics* 2002; **16**: 482–489.
11. Gestel T, Suykens J, Lanckriet G, Lambrechts A, De Moor B, Vandewalle J. Bayesian framework for least-squares support vector machine classifiers, Gaussian processes, and Kernel Fisher discriminant analysis. *Neural Comput*. 2002; **14**: 1115–1147.
12. Thissen U, Ustun B, Melssen W, Buydens L. Multivariate calibration with least-squares support vector machines. *Anal. Chem*. 2004; **76**: 3099–3105.
13. Chauchard F, Cogdill R, Roussel S, Roger J, Bellon-Maurel V. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. *Chemometrics Intell. Lab. Syst*. 2004; **71**: 141–150.
14. Suykens J, Vandewalle J, De Moor B. Optimal control by least squares support vector machines. *Neural Netw*. 2001; **14**: 23–35.
15. Muller K, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw*. 2001; **12**(2): 181–202.
16. Scholkopf B, Smola A, Muller K. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput*. 1998; **10**: 1299–1319.
17. Wu W, Massart D, Jong S. Kernel-PCA algorithms for wide data Part II: Fast cross-validation and application in classification of NIR data. *Chemometrics Intell. Lab. Syst*. 1997; **37**: 271–280.
18. Bialkowski S. Species discrimination and quantitative estimation using incoherent linear optical signal processing of emission signals. *Anal. Chem*. 1986; **58**: 2563–2567.
19. Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem*. 1986; **58**: 1167–1172.

20. Booksh K, Kowalski B. Theory of analytical chemistry. *Anal. Chem*. 1994; **66**(15): 782–791.
21. Lorber A. Net analyte signal calculation in multivariate calibration. *Anal. Chem*. 1997; **69**(8): 1620–1626.
22. Boelens H, Kok W, Noord O, Smilde A. Performance optimization of spectroscopic process analyzers. *Anal. Chem*. 2004; **76**(9): 2656–2663.
23. Prakash A, Stellman C, Booksh K. Optical regression: a method for improving quantitative precision of multivariate prediction with single channel spectrometers. *Chemometrics Intell. Lab. Syst*. 1999; **46**: 265–274.
24. Cassis L, Urbas A, Lodder R. Hyperspectral integrated computational imaging. *Anal. Bioanal. Chem*. 2005; **382**: 868–872.
25. Haibach F, Greer A, Schiza M, Priore R, Soyemi O, Myrick M. On-line reoptimization of filter designs for multivariate optical elements. *Appl. Optics*. 2003; **42**(10): 1833–1838.
26. Myrick M, Soyemi O, Li H, Zhang L. Spectral tolerance determination for multivariate optical element design. *Fresenius J. Anal. Chem*. 2001; **369**: 351–355.
27. Nelson M, Aust J, Dobrowolski J, Verly P, Myrick M. Multivariate optical computation for predictive spectroscopy. *Anal. Chem*. 1998; **70**: 73–82.
28. Soyemi O, Eastwood D, Zhang L, Li H, Karunamuni J, Gemperline P, Synowicki R, Myrick M. Design and Testing of a Multivariate Optical Element: The First Demonstration of Multivariate Optical Computing for Predictive Spectroscopy. *Anal. Chem*. 2001, **73**: 1069, 1079.
29. Weisstein E. Multiplex advantage—from Eric Weisstein's world of physics. Retrieved 5/01/05. http://scienceworld.wolfram.com/physics/MultiplexAdvantage.html/.
30. Hamilton S, Lodder R. Hyperspectral imaging technology for pharmaceutical analysis. *Proc. Soc. Photo-Optical Instrum. Eng*. 2002; **4626**: 136–147.
31. Nahorniak M, Booksh K. Optimizing the implementation of the PARAFAC method for near-real time calibration of excitation-emission fluorescence analysis. *J. Chemometrics*, 2003; **17**: 608–617.
32. Amanda L, Kah-Siew H, Robert A. Lodder, remote hyperspectral imaging of endolithic biofilms using a robotic probe. *Contact Context*: v1i1/planetprobe, 2002; **1**(1): 1–10.
33. Medendorp J, Lodder RA. Acoustic-resonance spectrometry as a process analytical technology for rapid and accurate tablet identification. Pharmscitech. In press.

