# Hyperspectral NIR Image Analysis

## Data Exploration, Correction, and Regression

James E. Burger
*Unit of Biomass Technology and Chemistry*
*Umeå*

**Doctoral Thesis**
**Swedish University of Agricultural Sciences**
**Umeå 2006**

**Acta Universitatis Agriculturae Sueciae**
2006: 60

ISSN 1652-6880
ISBN 91-576-7109-5

Printed by: Arkitektkopia, Umeå, Sweden. 2006.

# Abstract

Burger J. 2006. Hyperspectral NIR Image Analysis: Data Exploration, Correction, and Regression. Doctoral Dissertation.
ISSN 1652-6880, ISBN 91-576-7109-5.

Hyperspectral images add a new dimension to the field of spectroscopy, specifically spatial resolution. In addition to the identification and quantification of bulk constituents provided by integrating spectrometers, hyperspectral images provide a means of accurately quantifying and locating constituent variation within the field of view of the camera.

Hyperspectral images provide a massive quantity of data, and as with NIR spectroscopy, multivariate chemometrics tools must be utilized to appropriately extract accurate information. This thesis looked at techniques to clean and modify or condition the raw spectral data to improve the prediction results of regression techniques such as PLS. It was found that extra diagnostic tools for regression models could be based on image data. A new metric based on a combination of prediction bias and variance was proposed for determining the number of latent variables.

Data set conditioning was based on several approaches. Sets of standard reference materials were used to improve conversion of data counts into percent reflectance units and to provide instrument standardization. A multi-step approach to outlier detection was formulated that incorporated thresholding tests for excessive data values, combined with tests based on Euclidean distance measurements and angle cosines between spectra. Finally, various spectral pretreatments or filters were considered to complete the spectral cleaning and modification process.

Results from the application of multivariate analysis techniques to this optimally conditioned data were presented. Data visualization tools included histograms and spatial mapping of constituent concentration predictions, colorization of score plots, and false color image presentations of combinations of score images or prediction maps.

The use of these data exploration, correction, and regression tools was demonstrated by the systematic analysis of increasingly complex data samples. Carefully designed laboratory samples were used to examine the theoretical limitations of prediction of chemical content and correction for physical properties including the dependencies of diffuse light scattering effects on particle size. Sample sets of cheese and wood pellets were used to demonstrate the overall utility of proper data conditioning in the application of hyperspectral NIR imaging to more difficult real-world problems.

*Keywords*: image standardization, reflectance transformations, spectral preprocessing, scatter correction, multivariate image regression, outlier detection and correction

*Author's address*: James E. Burger, Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, SE-904 03 Umeå, Sweden. Jim.Burger@btk.slu.se

to

dad and jeff

# Contents

# Appendix

## List of original papers

This thesis is based on the following papers which will be referred to in the text by their corresponding Roman numerals:

I. Burger, J. & Geladi, P. 2005. Hyperspectral NIR image regression part I: Calibration and correction. *Journal of Chemometrics* **19**, 355.

II. Burger, J. & Geladi, P. 2006a. Hyperspectral NIR image regression part II: Preprocessing diagnostics. *Journal of Chemometrics* 2006. (in print)

III. Burger, J. & Geladi, P. 2006b. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst*, 2006. (in print)

IV. Burger, J. & Geladi, P. 2006c. Spectral pre-treatments of hyperspectral NIR images: analysis of diffuse reflectance scattering. *Journal of Near Infrared Spectroscopy* (submitted)

Papers I, II, and III are reproduced by kind permission of the journals concerned.

## Related work

The author also contributed to the following publications:

1. Burger, J. & Geladi, P. 2006d. *Chemometric Tools for Exploration of Hyperspectral Images*. In: Near Infrared Spectroscopy: Proceedings of the 12th International Conference. NIR Publications, Chichester, UK. (in print)

2. Rodionova, O., Houmøller, L., Pomerantsev, A., Geladi, P., Burger, J., Dorofeyev, V., & Arzamastsev, A. 2005. NIR spectrometry for counterfeit drug detection – A feasibility study. *Analytica Chimica Acta* **549**, 151.

3. Geladi, P., Burger, J., & Lestander, T. 2004. Hyperspectral imaging: calibration problems and solutions. *Chemometrics and Intelligent Laboratory Systems* **72**, 209.

4. Geladi, P., Sethson, B., Nyström, J., Lillhonga, T., Lestander, T., & Burger, J. 2004. Chemometrics in Spectroscopy. Part 2. Examples *Spectrochimica Acta B* **59**, 1347.

# Nomenclature

| | |
|---|---|
| x | Single scalar value |
| **x** | Vector of scalar values |
| **X** | Two-dimensional matrix of scalar values |
| $\underline{\mathbf{X}}$ | Three-dimensional array of scalar values |
| $\mathbf{X}^T$ | The transpose of matrix **X** |

# Abbreviations

| | |
|---|---|
| EDA | Exploratory Data Analysis |
| JIMIA | Java Implementation of Multivariate Image Analysis |
| LCTF | Liquid Crystal Tunable Filter |
| MIA | Multivariate Image Analysis |
| MIR | Multivariate Image Regression |
| MSC | Multiplicative Scatter Correction |
| NIR | Near Infrared |
| PCA | Principal Component Analysis |
| PLS | Partial Least Squares |
| PMSC | Piecewise Multiplicative Scatter Correction |
| PRESS | Predicted Residual Error Sum Squares |
| PTFE | polytetrafluoroethylene |
| ROI | Region of Interest |
| RMSEC | Root Mean Square Error of Calibration |
| RMSECV | Root Mean Square Error of Cross Validation |
| RMSEP | Root Mean Square Error of Prediction |
| SNV | Standard Normal Variate |
| SRM | Standard Reference Material |
| SVD | Singular Value Decomposition |

# 1.0 Introduction

## 1.1 Hyperspectral images

Chemical analysis is often undertaken to identify or quantify individual chemical constituents contained within a sample mixture, for example the quantity of sugar or fat in food, or caffeine in coffee. Since laboratory instrumentation provides only signals based on measurable physical properties, a transformation process must be used to convert instrument signals into the primary properties of interest. Spectrometry is often used for this purpose based on the fundamental principle known as the Beer-Lambert law that permits the determination of constituent concentrations based on the absorbance of light. In this case a calibration model based on measurements acquired from a set of training samples containing known constituent concentrations is determined and used for the transformation process.

Spectroscopic detectors have historically been based on an integrating principle to improve signal to noise ratios. That is, a small spot or sample area is measured, sometimes multiple times, and the spectral intensities throughout the sample area are added together. This provides a bulk sample measurement which is applicable to homogenous materials. But if the sample is inhomogeneous different measured values will be obtained depending on the sampling location. Unless the sample measurement is repeated many times in a systematic way, this inhomogeneity will never be known. Worse yet, a single measurement may be erroneously used to represent the bulk value of the entire sample.

Recent advances in spectroscopic detector technologies have enabled spectral as well as spatial information to be recorded simultaneously. This results in creation of three dimensional spectroscopic images where two dimensions provide spatial information and one dimension provides spectral information. This three dimensional dataset of spectral intensities is termed a *hyperspectral image*, or *hypercube*, and adds an important aspect to chemical analysis. In addition to the identification and quantification of chemical constituents, their location or spatial distribution can be examined as well.

Figure 1 and Figure 2 present the historical development of hyperspectral images from an imaging perspective. Black and white photography was first used to create two dimensional images by compressing a range of colors measured throughout the image area into a set of gray values representing the average spectral intensity at each location in the image. By adding specific color filters to the camera or film, images could be acquired which represented narrow bands of the color spectrum. This is exemplified today in the typical color digital photograph, comprised of a set of three separate grayscale images representing red, green, and blue color bands as indicated in Figure 1. Such an image is said to have three channels of information.

Additional channels can be added by using additional sets of color filters depicted in Figure 2a. This technique was first popularized in the field of remote sensing where images of the earth's surface were acquired from spacecraft,

Figure 1 A typical digital color picture is a composite of three image channels representing red, green, and blue intensities.

satellites or airplanes. The Apollo 9 mission (1968) included four Hasselblad cameras aimed at the same target, while the Skylab missions (1973 - 1974) included a six camera multi-spectral system. The Landsat I satellite launched in 1972 contained a multispectral scanner which permitted the acquisition of red, green, and two infrared channels. (NASA, 2006) Similar multi-channel images have been used in microscopy where a filter wheel or similar device is used to acquire sets of images representing different color bands (Schotton, 1993). Images with four or more channels are called *multi-spectral images*.

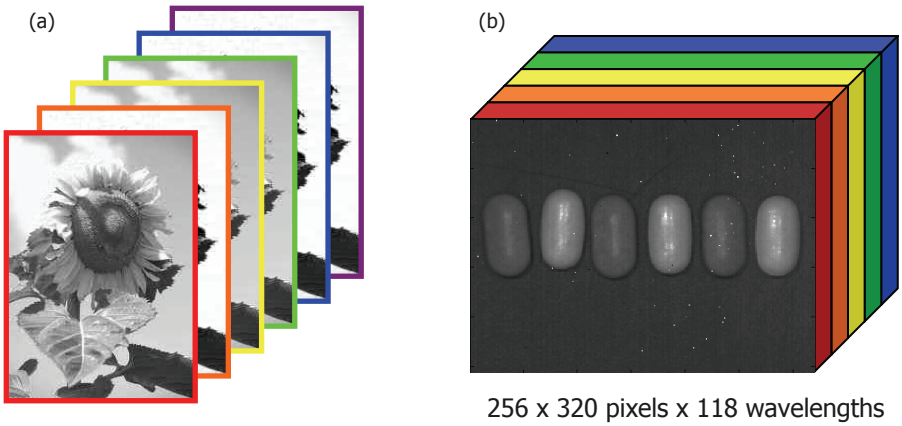


Figure 2 (a) Multi-spectral images contain channels from additional color bands obtained by use of 4 or more color filters. (b) A typical hyperspectral image containing 118 image channels and over 80,000 spectra.

When many different wavelength channels or bands are obtained, these images are termed *hyperspectral images*. There is no clearly defined number of channels

that differentiates multi-spectral from hyperspectral, but hyperspectral images typically contain over 100 wavelength channels. Figure 2b represents a typical hyperspectral image acquired from a commercial NIR imaging system used in this thesis. Hyperspectral images typically contain tens of thousands of spectra (one spectrum at each pixel location) and hundreds of channels or wavelength bands. This abundance of data within a hyperspectral image provides both challenges and opportunities for analysis. Which spectra are important to look at? If a spatial image is desired, which channel should be selected? Data mining tools must be used to efficiently extract meaningful information from this massive amount of data, and they must be used correctly.

## 1.2 Multivariate Image Analysis (MIA)

The theoretical aspect of extracting chemical information from hyperspectral images was first presented by Geladi, Wold, & Esbensen (1986). Although a hyperspectral image is a three dimensional data structure, it can be unfolded to produce a very large matrix of spectra. Such a matrix is amenable to the ordinary chemometric tools used in the analysis of spectral data. Fundamental work in the application of chemometric tools began with Multivariate Image Analysis (MIA) discussions on the application of Principal Component Analysis (PCA) (Geladi *et al.*, 1988) and the presentation of an analysis strategy (Esbensen & Geladi, 1989). Additional essential theoretical work in Multivariate Image Regression (MIR) included the application of Principal Component Regression (PCR) (Geladi & Esbensen, 1991; Esbensen, Geladi, & Grahn, 1992) and Partial Least Squares (PLS) (Van den Broek *et al.*, 1996; Lied, Geladi, & Esbensen, 2000; Lied & Esbensen, 2001) Geladi (1995) also introduced the concept of image sampling and local modeling specific to hyperspectral images. The image analysis toolbox has grown to include the application of Independent Component Analysis (ICP) (Chiang, Chang, & Ginsberg, 2000), wavelet based classification (Antonelli *et al*., 2004; Vogt, Banerji & Booksh, 2004) and support vector machines (Pierna *et al*., 2004).

The spatial scaling represented in hyperspectral images ranges from satellite images with a field of view of square miles, to the millimeter x millimeter images acquired through a microscope. This range in scaling can be found in hyperspectral imaging applications ranging from the discrimination of crops and weeds (Borregaard *et al*., 2000) to pharmaceutical tablets (Zhang, Henson, & Sekulic, 2005).

The range in sample types is equally broad. Research topics range from the analysis of painted chinaware (Geladi, Swerts, & Lindgren, 1994), works of art (Baronti *et al*., 1997; Attas *et al*., 2003; Bacci *et al*., 2005) and archaeology (Marengo *et al*., 2005), to process monitoring of plastic types (Van den Broek *et al*., 1996), snack foods (Yu & MacGregor, 2003), poultry carcass contamination (Park *et al*., 2004), or industrial boiler flames (Yu & MacGregor, 2004).

A complete overview of near infrared imaging principles and instrumentation is provided by Tran (2003). While the majority of reported research is based on visible or near infrared hyperspectral images, other sources of radiation have been utilized: Nuclear Magnetic Resonance (NMR) (Grahn *et al*., 1989a; Grahn *et al*., 1989b; Grahn & Saaf, 1992) X-ray Photoemission Spectroscopy (XPS) (Artyushkova & Fulghum, 2001, 2002), FT-IR (Winson *et al*., 1997) and Raman spectroscopy (Zhang, Henson, & Sekulic, 2005).

## 1.3 Terminology

Acquisition, exploration, and analysis of hyperspectral images are complex and multi-disciplinary tasks. It is important to clarify terminology used in this thesis to avoid miss-use and confusion. The terms *hyperspectral image* and *hypercube* are used interchangeably and denote a three dimensional data structure containing two *spatial* axes and one *wavelength* axis. A single *scalar* data element in a hypercube is a *voxel*, but within the context of a two dimensional image from a single wavelength channel it is termed a *pixel*. The set of pixels at the same location in all wavelength channels is a *vector*, traditionally called a *spectrum*.

This thesis deals extensively with the manipulations of numerical data. It is important to distinguish between the actions of transformation, standardization, and calibration. Data *transformation* involves the conversion of data from one set of units to another: e.g. raw instrument data may be transformed to reflectance; reflectance data may be transformed to absorbance. *Standardization* is a correction process used to move data towards a targeted value: e.g. internal standards may be used for instrument standardization; multiplicative scatter correction (MSC) moves data towards a targeted spectrum. *Calibration* is used in the context of chemometrics: calibration data and calibration models are used to predict values of dependent variables from related independent variables.

The use of these terms is not always clear. Some processes may cross these boundaries: the derivative of a spectrum may be considered either a transformation process changing units, or a correction or standardization process removing a baseline offset. Regression modeling which is considered a calibration technique, may be used to compute reflectance transformations or instrument standardizations. Some terms have multiple meanings: for example a *sample* or *sample point* may refer to a laboratory sample, a spectrum from a hypercube, or a reflectance value at one wavelength in a spectrum. Hopefully the intended meanings will be clear from their contextual use.

## 1.4 Research objectives

One of the general questions addressed in this thesis is: How can existing chemometric analysis tools can be used or enhanced to effectively and efficiently explore both the qualitative and quantitative information contained within a hyperspectral image? The reported use of chemometric tools on hyperspectral images has been limited primarily to tasks of exploratory data analysis and data compression using PCA, or discriminant analysis of pixels, the latter being

achieved with PLS calibration models based on masks of image pixels with known classifications. PLS is routinely used to quantify chemical constituents in samples based on spectroscopic data. A slightly more specific research question in this thesis is: What are the implications of applying PLS to hyperspectral images for quantitative purposes? Various aspects of the application and development of univariate and multivariate chemometric tools and their specific extension to hyperspectral images is discussed further in section 3 of this thesis, Chemometrics.

A second general focus of research for this thesis was on how to objectively determine the quality of data within hyperspectral images, as well as to explore methodologies for improving the quality of such data. Sometimes experimental data must be screened for faulty values or pretreated to remove undesirable contributions of signal noise or offset due to instrumentation or environmental effects. This thesis examines the causes, effects, and possible corrections of such deviations in hyperspectral images. Specifically, what can be done to promote instrument standardization in hyperspectral imaging systems? How can instrumentation failures be detected, and can they be corrected? These questions are addressed in two sections. Section 4, Hyperspectral Image Transformation and Standardization, describes the use of external and internal calibration reference materials to standardize each hyperspectral image. This permits the direct comparison of images taken sequentially in a given day, or perhaps months apart, and facilitates correct matching of spectra to spectral libraries. Hyperspectral imaging systems additionally employ camera devices which may include pixels that routinely or randomly produce faulty data. How should these outlier points be detected? Several approaches to this are described in section 5, Outlier Detection.

Physical properties of the samples or sample delivery system may produce additional perturbations to the spectra measured. The effect of spectral pretreatment techniques used in other forms of spectroscopy and how they can be extended to hyperspectral images is discussed in section 6, Spectral Preprocessing.

Visual examination of objects involves not only spectroscopic or color analysis, but also spatial information such as texture, object shape, or even contextual information defining the object's environment. An interactive exploratory analysis tool was developed to enable simultaneous examination of both spectral and spatial properties of hyperspectral images. This tool is explained in section 7, Java Implementation of Multivariate Image Analysis (JIMIA).

The success of all of these topics is considered in the final sections. Concluding results for the application of these various analysis techniques are summarized in section 8, Applications. Concluding thoughts and extensions beyond the scope of this thesis are described in section 9, Conclusions, and section 10, Future Directions.

Much of the discussions provided in this thesis can be applied in general to all hyperspectral images. However, the research experimentation performed in support of this thesis utilized exclusively near infrared (NIR) spectroscopy on

samples within a laboratory environment. Consequently, discussions will at times remain focused on these environs.

# 2.0 NIR Instrumentation

## 2.1 NIR Spectroscopy

The near infrared region of the spectrum is usually defined as the range 780 – 2500 nm. This region is located between the infrared region above 2500 nm, and the visible region below 780 nm. The first recorded NIR measurements were made by Herschel (1800) while exploring the heat energy from the sun. NIR spectra of several organic liquids were measured photographically and reported by Abney & Festing (1881). Functional group peak assignments began when Coblenz (1905) measured compounds between 800 and 2800 nm, and attributed the bands observed to C-H bonds. Additional peak assignments were made at the University of California at Los Angeles and Johns Hopkins University 1922 – 1929 (Ellis 1929), by the US Bureau of Standards between 1930 and 1945, Murray in the UK during the 1980's and Barton at the USDA in the 1980's (Barton 2004). Norris & Butler (1961) and Norris (1962) were also instrumental in promoting NIR as an acceptable analytical technique while working at the U.S. Department of Agriculture, Agricultural Research Service. In general the peaks observed in the NIR region are due to overtone vibrations between hydrogen and carbon, oxygen, or nitrogen. These peaks are very broad and overlapped, and lack fine structure. Spectra from different constituents can appear very similar and difficult to resolve. For this reason the NIR region of the spectrum was generally ignored until the advent of the personal computer and the application of multivariate statistics and chemometrics that allowed mathematical rather than physical resolution of the overlapping spectra. The general philosophy and practice of NIR spectroscopy is well described by Murray (2004) and Osborn, Fearn, & Hindle (1993).

A second reason for the growth in popularity of NIR has been the development of new detector systems. Communication research led to the initial development of photoelectric detectors during the Second World War. These devices were first used commercially for NIR detection with the release of the Cary 14 NIR spectrometer (Kaye 1954). Development of Si, PbS, and InGaAs detectors has progressed to allow operation at room temperature, extended wavelength response, and fabrication into one and two dimensional arrays. The current state of the art detectors used for hyperspectral NIR imaging employ two dimensional InGaAs detectors with 512 x 640 elements and provide a usable wavelength response range of 900 – 1700 nm or an extended range of 1100 – 2500 nm. These focal plane cameras are thermoelectrically cooled and provide full images at up to 90 frames per second. (XenICs NV, Leuven, Belgium).

## 2.2 Hyperspectral NIR camera configurations

An NIR spectrophotometer instrument consists of a light source, a monochrometer or filter system to disperse the light into wavelength bands, and a detection system including both a detector and any necessary controlling software (Stark & Luchter, 2004). Tungsten halogen filament or xenon gas plasma lamps are the most common sources used for broad spectral NIR illumination. Light emitting diodes and tunable lasers may also be used for illumination with specific wavelength bands. For broad spectral sources, selection of wavelength bands can be based on specific bandpass filters based on simple interference filters, liquid crystal tunable filters (LCTF), or acousto-optic tunable filters (AOTF), or the spectral energy may be dispersed by a grating device or a prism-grating-prism (PGP) filter (Aikio, 2001; SpecIm, 2006). Scanning interferometers can also be used to acquire NIR spectra from a single spot.

A spectrometer camera designed for acquisition of hyperspectral images has all the hardware components listed above for acquisition of spectral information plus any additional hardware or modifications necessary to allow acquisition of reproducible spatial information. The spatial information is acquired by measurement directly through the spectrometer optics, by controlled positioning of the sample, or both. Three basic camera configurations are used which differ in their need for positioning or movement of the sample and their geometry or detector orientation relative to the sample. The configuration names are based on the type of spatial information acquired: point scan, line scan, or plane scan.

### *2.2.1 Point scan*

The interferometer camera configuration shown in Figure 3 is perhaps the simplest approach. A complete spectrum is measured from the diffuse reflectance or transmitted light at a single spot on the sample. The sample is then repositioned before obtaining a new spectrum. By rastering the sample in two spatial dimensions normal to the sample transmission or reflectance radiation beam, a complete hyperspectral image can be acquired. Instrument calibration for wavelength and reflectance needs to be done only once before beginning the image acquisition. However, any drift during the total acquisition period is difficult to detect without recalibration. This system provides very stable high resolution spectra; however, the sample positioning is very time consuming and places high demands on repositioning hardware to ensure repeatability. Image registration of the different image wavelength channels is only an issue if something in the sample moves within the spectral scanning period. The spatial size dimensions of the hyperspectral image are limited only by the sample positioning hardware. Continuous operation permits extended sample sizes or resampling at specific spatial locations. This configuration was first patented by Potter in 1972 working at NASA and implemented for atmospheric and astronomical application in 1979 by Potter & Huppi at the Stewart Radiance Laboratory. (Miseo & Wright, 2003) It is primarily used in the chemical laboratory at the microscopic level (Sahlin & Peppas, 1997) and is commercially

available from Digilab (Randolph, MA, USA), Bruker Optics (Billerica, MA, USA), Varian (Palo Alto, CA, USA), and Thermo Nicolet (Madison, WI, USA)

### *2.2.2 Line scan*

The second camera configuration approach indicated in Figure 4 is termed a linescan or pushbroom configuration that uses a two dimensional detector perpendicular to the surface of the sample. The diffuse reflectance from the sample is passed through a narrow slit and spectrally dispersed across the detector.

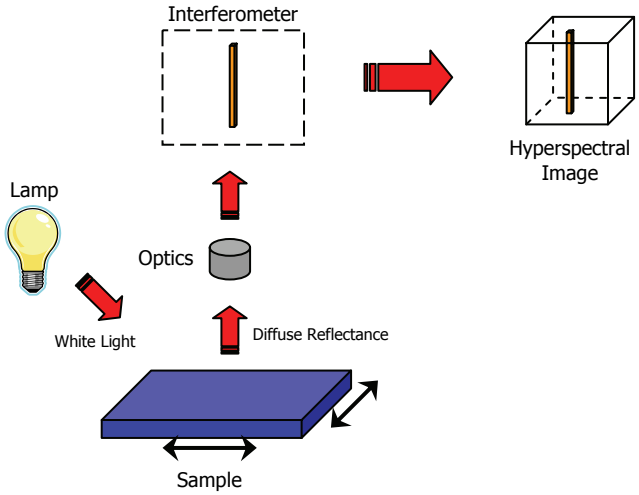


Figure 3 An interferometer measures a complete spectrum at a single spot. A hyperspectral image is created by rastering across the surface in two spatial dimensions.

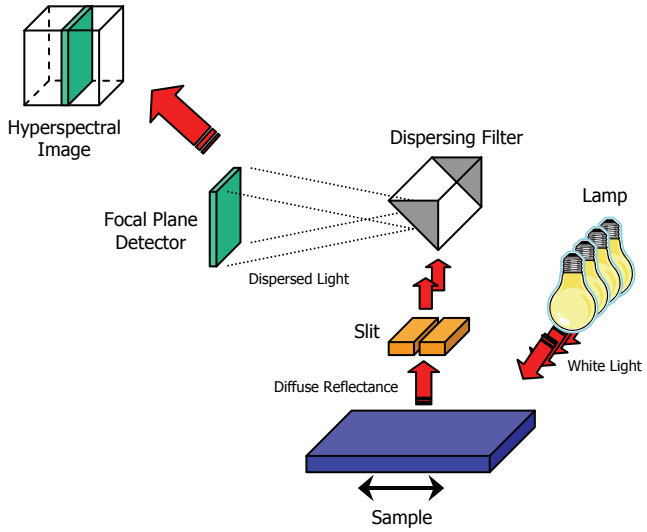


Figure 4 The linescan configuration uses a two dimensional detector perpendicular to the surface of the sample to capture the complete spectra from a set of points across the surface of the sample. Hyperspectral images are created from collecting sets of these matrices while repositioning the sample in a direction perpendicular to the sample scan line.

Hyperspectral images are created by collecting sets of these matrices while repositioning the sample in a direction perpendicular to the sample scan line. As with the rastering technique, the camera can be operated in a continuous mode permitting measurement of samples with an unlimited size in one axis. A two dimensional calibration is necessary to account for variations in lighting intensities across the spatial scanning line and wavelength dependent response variations. Without recalibration any instrumental variation across the scanning line or spectral axis is difficult to distinguish from sample variation. General instrument variations can be monitored by including reference standards at the edge of the field of view. Since no filter change is necessary, the speed of image acquisition is limited only by camera read out speeds. Commercial instruments produced by SpecIm (Oulu, FI) and Norsk Elektro Optikk AS, (Lørenskog, NO) are available with frame rates of 90 Hz or higher with 256 x 320 pixel resolution InGaAs detectors. This speed allows images to be acquired in a matter of seconds. This configuration is also amenable to continuous operation for online monitoring of process streams.

### *2.2.3 Plane scan*

The third configuration displayed in Figure 5 positions the detector in a plane parallel to the surface of the sample. All spectrometer components including the sample remain fixed in position relative to the detector, with the possible exception of the filter assembly. Interference filters must be rotated into position for each image slice. For LCTF or AOTF filters this filter change is done electronically. The lack of moving parts can lead to a lower background noise level; however, because of significant instrument settling times due to filter changes any movement within the sample itself will cause image registration problems. Lengthy image acquisition times can also be an issue for biological samples which may be sensitive to heating caused by the continuous illumination from NIR source lamps. The LCTF based instrument used in this thesis had image acquisition times ranging between three and six minutes that initially caused sample separation and melting problems. A small box fan positioned to blow cool air across the sample reduced the severity of this problem.

Since light illumination intensities may vary across both spatial axes of the sample area, instrument calibration must be performed at all pixel (spatial) locations and at all wavelengths. Additional standardization to account for instrumentation instabilities may be performed by including standard reference materials within the sample field of view. These calibration and standardization techniques are discussed in section 4 of the thesis.

Although detector materials such as InGaAs have been extended to cover broader ranges of the NIR spectrum, the quantum efficiency varies considerably. To account for this variation, it would be possible to vary the signal acquisition or integration time as a function of wavelength with the configuration depicted in Figure 5, thereby increasing the signal to noise ratio in the areas with lower efficiency. At present this is not done. The commercial instrument used in this thesis requires the same integration time for all wavelengths.

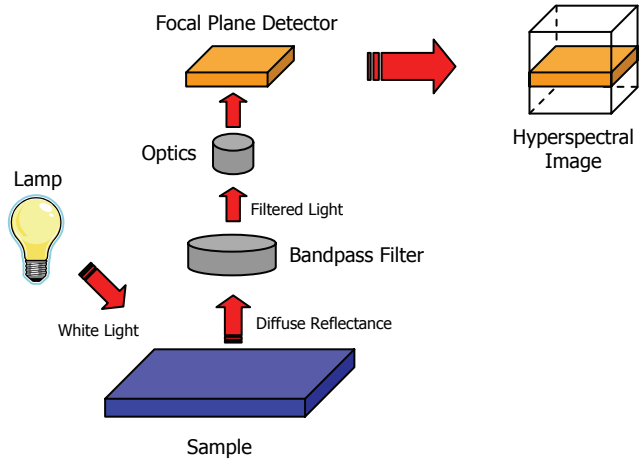


Figure 5 A bandpass camera captures a complete image scene slice one wavelength band at a time. Sets of these image slices are used to create a complete hyperspectral image. The sample remains fixed in position relative to the camera.

## 2.3 Research instrumentation

The hyperspectral images used in Papers I – IV in this thesis were all acquired from a MatrixNIR imaging system from Spectral Dimensions, Inc. (Olney, MD, USA). Images acquired from this instrument have a fixed spatial resolution of 256 x 320 pixels with up to 128 wavelength channels. This instrument is capable of providing spectra ranging between 900 and 1700 nm; however, because of high relative noise levels at either end of the spectrum, a working range of 960 to 1662 nm was typically used. A six nm resolution provided images with 118 wavelength channels. Images were generally measured with a 64ms signal integration time, averaged from 10 to 16 replicate scans.

For comparative purposes, some of the samples in this thesis were also measured with integrating NIR spectrometers. Two instruments were used for this. A Foss NIRSytems 5000 (NIRSystems, Hillerød, DK) scanning grating monochromator with a PbS detector provided spectra with a wavelength range of 1100-2498 nm with 2 nm resolution. Each spectrum was averaged from 32 scans. Samples were contained in glass covered sample holders which spun during signal acquisition. These same sample holders were also used for hyperspectral imaging

Spectra were also acquired from a Bruker Matrix FT-NIR instrument, (Bruker Optics, Billerica, MA, USA) equipped with a fiber-optic sampling probe. This instrument uses an InGaAs detector with a working range of 4000-12000 $cm^{-1}$. Spectra were collected at 4 $cm^{-1}$ resolution, but for instrument comparison purposes were displayed in wavelength units, 833 – 2500 nm. Each spectrum was an average of 32 scans. The probe was inserted directly into containers of powdered samples, or positioned directly perpendicular to the surface of solid samples.

# 3.0 Chemometrics

Chemometrics is a relatively new branch of science with several official and unofficial definitions:

> ***Chemometrics*** *is the chemical discipline that uses mathematical, statistical, and other methods employing formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analyzing chemical data.* (Massart *et al*., 1988)

> ***Chemometrics*** *is the discipline concerned with the application of statistics and mathematical methods, as well as those methods based on mathematical logic, to chemistry.* (Brown, 1990)

> ***Chemometrics*** *is what chemometricians do. Chemometricians are people who drink beer and steal ideas from statisticians.* (WSC-3, 2004)

The term *chemometrics* was introduced in the early 1970's by Bruce Kowalski and Svante Wold. With the introduction and widespread use of computers in the chemistry laboratory for both instrumentation control and data acquisition, it became apparent that new tools needed to be invented (or stolen) to make sense of the growing mass of chemical data. Statistical multivariate analysis had previously been applied in fields such as psychometrics, econometrics, and biometrics (Geladi & Esbensen, 1990; Geladi, 2003). Extending these techniques to chemistry was natural. Hence, chemometrics was born.

Borrowing a term from computer science, chemometrics might also be thought of as *data mining*. Algorithms are needed to compress and model numerical data in a systematic yet meaningful way to extract usable information. Presentation tools are also needed to help visualize this information buried within the tremendous amounts of data.

One of the fundamental principles of chemometrics can be loosely termed *get to know your data*; before any attempts at modeling or prediction, experimental data sets should always be thoroughly examined. Basic statistical descriptors and population estimates of data sets such as the minimum, maximum, mean, median, and standard deviation of the data values should be determined. A visual inspection of both tabular and plotted data can also help to identify outliers, clustering of values, data trends, and an indication of underlying noise structures. This is a simple process when looking at univariate measurements such as temperature or pH of a few samples. But a single hyperspectral image with 256 x 320 pixels and 128 wavelength channels contains over twenty million data values, data which can be acquired in less than 5 seconds. How does one quickly but effectively explore such a massive amount of data?

Clearly the simple plotting and exploratory analysis tools of univariate and multivariate statistical analysis are inadequate and need to be enhanced and adapted for these large data sets. Regression techniques such as Partial Least Squares (PLS) are often applied to spectral data to build models useful for making chemical classification or quantification predictions. But when such models are applied to a single hyperspectral image as described above, over eighty thousand predictions will be produced. How do you efficiently present these results for interpretation? While the shear volume of data contained in hyperspectral images may appear overwhelming and demand powerful computational resources, it also provides new opportunities to enhance our understanding of relevant chemical information. Extending the chemometrics toolbox to facilitate efficient computations and enhance visualization of results will be further explored after the following theoretical discussions.

## 3.1 Data structures

The mathematical based term *order* of a dataset was introduced by Sanchez & Kowalski (1988a; 1988b) to describe the true dimensionality of a dataset. A pH meter or thermometer is zero order, producing a single data value with each measurement. A UV spectrometer is said to be first order: each measured spectrum is a vector of numbers indicating absorbance values obtained at different wavelengths. A hybrid instrument such as a gas chromatograph – mass spectrometer (GC-MS) produces second order data matrices containing counts of ions for different mass numbers, as a function of time. A data set collected from a sensory panel of judges characterizing a collection of wines with a series of sensory taste tests is third order: judge vs. wine vs. test, and so on. The term *dimension*, often confused with order, has several meanings. The number of variables in a given order may be called dimensions e.g., the number of judges in the wine tasting dataset. Dimension may also refer to the number of orthogonal axes in a dataset e.g., we see objects in a three dimensional world. Time is sometimes considered a forth dimension. This latter definition of dimension will be used throughout this thesis. The dimension and the order of a dataset may not necessarily be the same.

What is the order of a hyperspectral image? Consider a hyperspectral image $\underline{\mathbf{X}}$ with I rows of pixels, J columns of pixels, and K wavelengths, as depicted in Figure 6. This dataset appears to be three dimensional, but this data *cube* effectively contains a set of I•J spectra, or alternatively K gray-scale single channel images at different wavelengths. Is a hyperspectral image third order, or perhaps only second order? Consider first a two dimensional data set containing a matrix of spectra acquired over time or vs. distance from a fixed point; this is clearly second order. (i.e. wavelength vs. time or wavelength vs. distance) Now combine a set of these two dimensional matrices, forming a three dimensional matrix of spectra acquired over time as a function of distance to a fixed point; this data set is third order with dimensions representing time, distance, and wavelength. But a hyperspectral image is in between; it has a three dimensional dataset, but the data units represent distance vs. distance vs. wavelength and are not truly third order, but perhaps two and a half. This is significant in that third

order analysis techniques such as PARAFAC and TUCKER (Smilde, Bro, & Geladi, 2005) are limited in their application to hyperspectral images. Higher order problem formulations of multivariate data is discussed further by Esbensen, Wold, & Geladi (1988).

Hyperspectral images can however be *unfolded* to create a two dimensional dataset to make use of second order analysis procedures. Figure 6 demonstrates this technique; sets of spectra are removed from a hyperspectral image one row of pixels at a time and arranged sequentially. This creates a very long matrix of spectra with a total size of number of pixels I • J vs. number of wavelengths K. By retaining the pixel location information of each spectrum, spectral prediction or processing results can be mapped back to a spatial location, creating an image map of results. This *back-folding* exploits the extra *half order* in the original hyperspectral image and demonstrates one of the primary advantages (spatial information) over other spectroscopic techniques based on sample area integration alone.

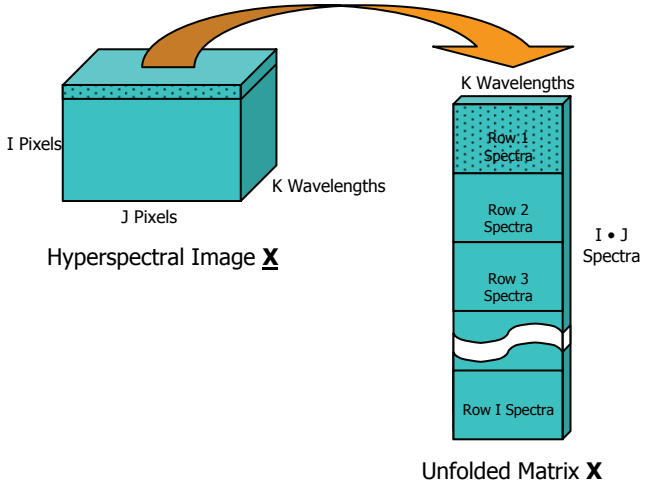


Figure 6 The three dimensional hyperspectral image **X** is unfolded into a two dimensional matrix **X**.

## 3.2 Principal Component Analysis (PCA)

One approach to reducing the data complexity in a hyperspectral image **X** is to first use multivariate data compression tools such as Principal Component Analysis (PCA) on the unfolded spectral data set **X.** This process separates signal from noise by projecting the data into a lower dimensional data space. PCA was first proposed by Pearson (1901) as "(a way of finding) lines and planes of closest fit to systems of points in the space." Axes in the original data space are rotated to align with the maximum signal variance. This process results in a bi-linear

decomposition as depicted in Figure 7. (Wold, Esbensen, & Geladi, 1987; Jackson, 1991; Malinowski, 2002) The column vectors contained in **P** are termed *loading* vectors, while those in **T** are called *score* vectors. The matrix **E** represents any *residual* noise not included in the **P** and **T** decomposition. A is the number of *latent variables* or components retained in the decomposition and ranges between 1 and K, the total number of independent variables. PCA finds the loading and score vectors in a decreasing order of importance: the first loading vector $\mathbf{p}_1$ is the loading vector which accounts for the most variance in the data structure. $\mathbf{p}_2$ is the next most significant component, etc. Data compression is achieved by selecting only the first A latent variables to represent the signal in **X**. The remaining A+1 to K latent variables characterize noise and are contained within the residual matrix **E**.

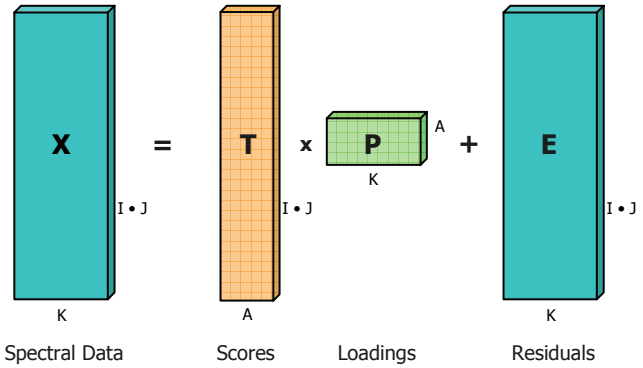


Figure 7 PCA is a bi-linear decomposition of data into scores, loadings, and residuals.

### *3.2.1 Computation considerations*

For numerical computation purposes PCA is routinely computed by a Singular Value Decomposition (SVD) algorithm first introduced by Beltrami (1873) and Jordan (1874). For a modern description see Golub & Reinsch (1970). Here **X** is decomposed into the product of matrices **U**, **S**, and **V**, where **S** is a diagonal matrix containing the K singular values of **X**, **V** contains the K eigenvectors of **X**, and **U** contains the K normalized score vectors. By definition **U** and **V** contain orthonormal column vectors and are ordered to reflect a decreasing sequence of singular values in **S**.

$$\mathbf{X} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{T} \qquad [1]$$

Because of the large numbers of spectra (I•J rows in **X**) and their possible co-linearity, problems with computation speed and numerical stability may occur in the SVD transformation of **X**. Geladi *et al.* (1988) showed that the SVD

transformation of both **X** and the cross product, covariance or correlation matrix $\mathbf{X}^T\mathbf{X}$ produced the same eigenvectors **V**. From Equation 1 it follows:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\,\mathbf{S}^T\,\mathbf{U}^T\,\mathbf{U}\,\mathbf{S}\,\mathbf{V}^T \qquad [2]$$

But since **U** contains orthonormal vectors, $\mathbf{U}^T\mathbf{U}$ equals the identity matrix and can be removed.

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\,\mathbf{S}^T\,\mathbf{S}\,\mathbf{V}^T \qquad [3]$$

Since S is a diagonal matrix with decreasing elements, the product matrix $\mathbf{S} = \mathbf{S}^T\mathbf{S}$ is as well. Therefore Equation 3 can be further simplified:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\,\mathbf{S}\,\mathbf{V}^T \qquad [4]$$

The SVD of the covariance matrix, Equation 5, has the exact same form as equation 4. Since both **V** and $\mathbf{V}_c$ contain orthonormal eigenvectors, and both **S** and $S_c$ contain diagonal elements with decreasing value, it follows that **V** and $\mathbf{V}_c$ must be identical.

$$\mathbf{X}^T\mathbf{X} = \mathbf{U}_c\,\mathbf{S}_c\,\mathbf{V}_c{}^T \qquad [5]$$

This is a significant contribution to the PCA of hyperspectral images. Basing the SVD on $\mathbf{X}^T\mathbf{X}$ is much faster since the cross product matrix dimensions are K x K which is considerably smaller. This matrix also lacks the collinearity problems of **X** making the SVD less prone to numerical instabilities. Once the loading vectors **V** are known the score vectors can easily be computed:

$$\mathbf{XV} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T\mathbf{V} = \mathbf{U}\,\mathbf{S} = \mathbf{T} \qquad [6]$$

### *3.2.2 Local modeling*

As will be demonstrated later, outlier data points with high leverage will greatly influence the orientation of the PCA loading vectors. Such data points must be identified and removed in order to enable PCA to find the true axes representing data variance. Optimal transformations will also be achieved by selecting only portions of the image that truly represent the sample of interest. This subset selection is termed local modeling. Figure 8 shows the difference between global PCA modeling (all spectra) and local PCA modeling of spectra selected by an image mask. The $\mathbf{t}_1$-$\mathbf{t}_2$ score plots have been colorized to indicate relative counts of points with identical pairs of score values, $t_{1i}$, $t_{2i}$ (see section 3.4.3).

This hyperspectral image is a view of four rectangular pieces of cheese surrounded by metal frames containing Spectralon internal standards. The global PCA loading vectors have become aligned with the strong variation in internal standard and frame spectra as well as outlier spectra, indicated by the long tail like structure and scattered points seen in (b). By limiting PCA modeling to only spectra from the cheese areas selected with the mask indicated in (c), well defined

clusters of the cheeses are obtained as seen in the score plot (d). In this local modeling example two of the cheeses could not be separated because they are almost identical in chemical composition.

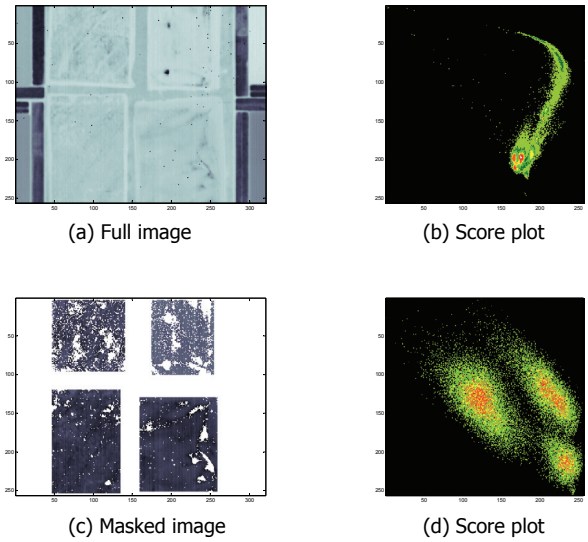


Figure 8 Global PCA modeling of the full image (a) results in a score plot (b) spanning the full image space. The well defined clusters in score plot (d) result from the local PCA model based on a subset of spectra indicated by the selection mask (c).

## 3.3 Partial Least Squares regression (PLS)

### *3.3.1 Regression*

The Beer-Lambert law expresses the linear relationship between spectral absorbance A and concentration c:

$$A = \varepsilon \text{ ł } c \qquad [7]$$

Here ł represents the path length of the sample. The molar absorbtivity ε is specific to each analyte and is a function of wavelength. Since the path length is typically constant for a given instrument it is often combined with ε and inverted, $b = \varepsilon^{-1} \text{ ł }^{-1}$, allowing Equation 1 to be written in an inverse form relating concentration y and spectral absorbance x:

$$y = x\, b \qquad [8]$$

For sets of complete spectra this can be written in a mean centered matrix form:

$$\mathbf{y} = \mathbf{X}\,\mathbf{b} + \mathbf{e} \qquad [9]$$

Where

**y** is a vector of mean centered concentrations
**X** is a matrix containing mean centered rows of spectra
**b** is a vector of constants relating the molar absorbtivity to wavelength
**e** represents the total measurement error

Partial least squares (PLS) is a multivariate regression technique which estimates **b** in such a way as to maximize the covariance between **y** and **X**. (Sjöström *et al.*, 1983; Frank, 1987; Höskuldsson, 1988). Latent variables or components are found which are aligned with the maximum variance of **X** as in PCA, but incorporate the covariance with **y** as well. These components are determined on the basis of decreasing significance, and choosing the correct number of latent variables A to include in the calibration model is essential for valid predictions. The weights of the first A chosen components are combined, resulting in the regression vector $\mathbf{b}_{hat}$ which can then be used to predict concentrations of analyte in new spectra:

$$\mathbf{y}_{hat} = \mathbf{X}\,\mathbf{b}_{hat} \qquad [10]$$

The residual error **r** is a bias measurement or difference between the predicted values $\mathbf{y}_{hat}$ and external reference values **y**, obtained from wet chemical analysis:

$$\mathbf{r} = \mathbf{y} - \mathbf{y}_{hat} \qquad [11]$$

### *3.3.2 Pseudorank diagnostics*

The number of latent variables A to include in the calibration model is called the *pseudorank*. If A is too small, termed *underfitting*, not enough signal is incorporated into the calibration model resulting in poor predictions. *Overfitting* occurs when A is too large and excessive noise is included in the calibration model. This results in a good fit of the spectra used for modeling, but poor predictions of additional spectra. (Martens & Næs, 1989; Beebe, Pell & Seasholtz, 1998; Næs *et al.*, 2002)

To assist with determining the pseudorank various approaches have been suggested for splitting the experimental data into a calibration or training set $\{\mathbf{y}_c\ \mathbf{X}_c\}$ and a test set $\{\mathbf{y}_t\ \mathbf{X}_t.\}$ *Leave-one-out cross validation* (LOO) is a special splitting where each sample is removed one at a time and predicted from a calibration model based on all the other samples (Hardy *et al.*, 1996). Diagnostic calibration model statistics can be computed from the sum of the squared residual errors for each of these approaches. The Root Mean Square Error of Calibration (RMSEC), Root Mean Square Error of Cross Validation (RMSECV), and Root Mean Square Error of Prediction (RMSEP) all indicate summary measurements of the errors or residuals in predictions. PRESS plots (Predicted Residual Error Sum Squares) can be made which graph these error diagnostics vs. the number of latent variables included in the model. A minimum value or breaking point in the plot may be observed when a sufficient number of variables have been included.

Hyperspectral images permit additional diagnostic information based on the predicted values and residual errors. First, since the spectra are obtained from specific pixel locations within an image, spatial coordinates can be used to map both the predicted and residual values, thereby forming prediction and residual images used for visual inspection. This technique is explained further in the next section.

Second, because of the large number of sample spectra within a hyperspectral image, accurate population statistic estimates can be obtained. It was observed that for the prediction results of an image region of interest (ROI) containing a single chemical constituent with known concentration, as the number of latent variables was increased, the prediction bias decreased because of better model fit, but at some point, the standard deviation began to increase because of overfitting. This can be seen in Figure 12 below.

A new diagnostic D-metric was introduced in Paper II based on pooling the average biases and the standard deviations of predictions from N different image ROIs:

$$BIAS_{pool} = [\ \mathbf{r}_{ave}^{T}\,\mathbf{r}_{ave} / N\ ]^{1/2} \qquad [12]$$

$$S_{pool} \approx [\ \mathbf{s}_{ave}^{T}\,\mathbf{s}_{ave} / N\ ]^{1/2} \qquad [13]$$

$$\text{D-metric} = [\ w_1\, BIAS_{pool}^{\ 2} + w_2\, S_{pool}^{\ 2}\ ]^{1/2} \qquad [14]$$

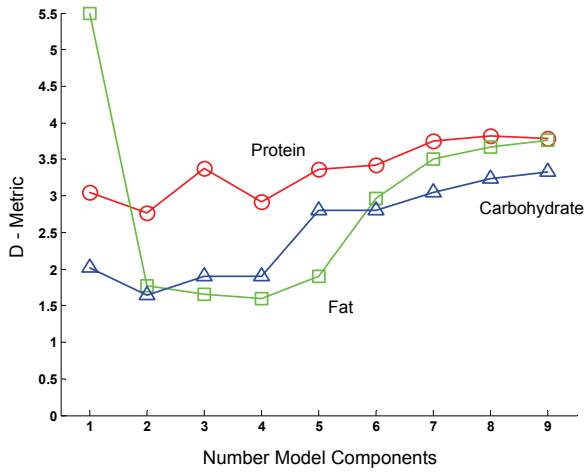


Figure 9 D-metric combining $BIAS_{pool}$ and $S_{pool}$ prediction errors in cheese. Protein (circle), fat (square), carbohydrate (triangle) vs. the number of components in the calibration model.

Here $\mathbf{r}_{ave}$ and $\mathbf{s}_{ave}$ represent the vectors containing the average bias and average standard deviation estimates from each of the N sample ROIs. The weights $w_1$ and $w_2$ were both set to the value 1.0 in Paper II however alternate weights could be used. Figure 9 (reprinted from Paper II) shows the D-metric vs. the number of model components for the prediction of fat, protein, and carbohydrate in cheese. In this case two PLS components was found to be an optimal number for all three ingredients. This new D-metric diagnostic combines both accuracy and precision variance measured throughout the hyperspectral image and was found to be quite robust to variations due to sample presentation errors.

### *3.3.3 Calibration dataset selection*

Since PLS models the variance in sample spectra, it is critical that spectral data sets be carefully selected to ensure maximum correlation between variance in independent variables (spectra) and dependent variables (concentration). The following two figures demonstrate problems caused by two sources of spectral variance.

It is well known that outlier points contribute high leverage in the computation of regression models. This is demonstrated in Figure 10 where the addition of a single outlier point greatly perturbs the regression model. For this reason it is essential that the entire set of hyperspectral image spectra be carefully examined and purged of all outliers. This process is examined further in a later section of this thesis.

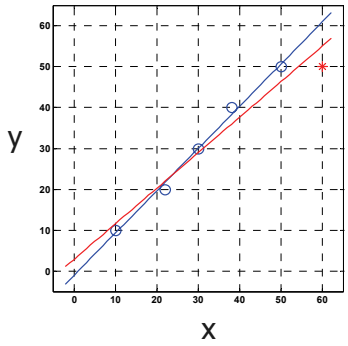


Figure 10 The outlier point * with high leverage perturbs the regression model. Two regression lines are indicated: the fit including the outlier point, and the cleaned dataset without the outlier (circle symbols).

Figure 11 demonstrates another possible danger in modeling hyperspectral images. Large amounts of noise exhibit leverage effects as well, even if normally distributed around expected values. Regression modeling based on mean values eliminates this extra variance. For this reason, it is recommended that hyperspectral image regression models be based on small sets of mean spectra from representative ROIs, rather than large sets of all spectra contained within the respective ROIs. A signal averaging effect is obtained by first computing the mean spectrum: thus, reducing the impact of spectral variance due to particle scattering or instrumentation effects.

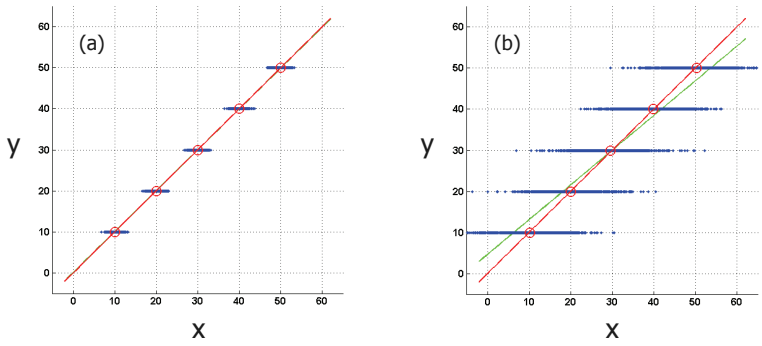


Figure 11 The effect of sample noise on linear regression of large sample sets: two sets of 5000 points containing 5 subsets of 1000 points each with N(y,1) distributions (a) and N(y,5) distributions (b). The circle symbols indicate the mean values of each of the 5 subsets. In both examples the regression line fit to the mean values passes through the circle symbols, whereas the regression line fitted to all 5000 points is indistinguishable in (a) but clearly perturbed in (b).

## 3.4 Exploratory analysis

### *3.4.1 Histograms*

What exploratory tools are available and how can they be fully utilized to examine the information content unique to hyperspectral images? Univariate Exploratory Data Analysis (EDA) tools such as histograms or box-whisker plots can certainly be applied, but ideally they should be applied in conjunction with other multivariate techniques such as PCA or PLS. For example, histograms of all the reflectance values for a particular wavelength channel can be examined for distribution profiles, mean, skewness, range, and especially outliers. But which wavelength channel should be chosen for examination? By first performing PCA, the same histogram diagnostics can be applied to scores, which better summarize the contributions of all channels.

Histograms of PLS predictions can be similarly displayed. Figure 12 indicates the distributions of the ~20000 predicted values for sugar concentration in an example hyperspectral image sample ROI. The distributions are presented sequentially as a function of the number of components included in the PLS model. It is clear to see that as the number of model components is increased, the prediction distributions move closer to the expected target value indicated by the vertical line. It can also be seen that the width of the distribution of predicted values increases, especially with 4 or more components. This trade-off in overall bias vs. standard deviation is what led to the computation of the D-metric described previously. Prediction histograms can also be examined for peak purity information. In this case the peaks appear unimodal with fairly Gaussian appearing distributions. Shoulder peaks or other anomalies can be useful in assessing sample uniformity. This information is unique to hyperspectral images

and is not available from spectra recorded from sample averaging or integrating spectrometers.

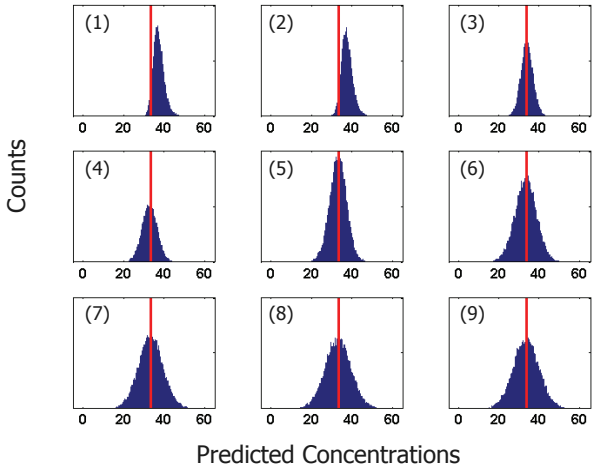


Figure 12 Sugar concentrations for 1-9 component PLS models. Summary histograms represent results from ~20000 spectra within a single image ROI. The vertical line indicates the target expected concentration of 33%.

### *3.4.2 Images*

The spatial information contained in a hyperspectral image **X** is totally ignored by PCA or PLS analysis. After unfolding, the order of the spectra contained in **X** can be totally randomized, producing identical PCA or PLS results. But by retaining the spatial coordinate information, PCA scores or PLS prediction results can be mapped spatially to create new images. Since PCA performs a significant data compression, the first few score images will indicate the greatest contrast in the spectral composition of image features. Gray scale score images can be further explored with more traditional processing techniques such as thresholding, edge detection, or application of two-dimensional transform functions applicable for feature recognition.

Both the prediction and residual values obtained from PLS can be similarly examined as gray scale images. Useful information regarding both the sample and instrumentation may be gained by examining these images created as a function of the number of latent variables. Systematically adding components may change the distribution or texture appearance of sample constituents. Adding additional components may also reveal other sources of signal variance such as lighting uniformity or other hardware stability issues.

In addition to the single channel gray scale images, combinations of images may be combined to provide false color mapping. Figure 13 from (Paper III) shows the spatial mapping for the predicted concentrations of sugar (a), citric acid (b), and salicylic acid (c) contained in a ternary mixture. The false color map image (d)

helps emphasize non-uniformity issues due to perhaps incomplete mixing of ingredients.

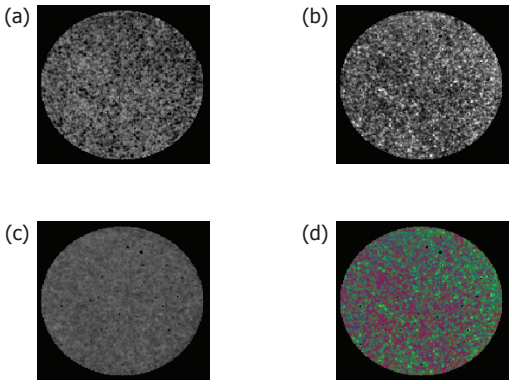


Figure 13 PLS prediction images for sugar (a), citric acid (b), and salicylic acid (c), can be overlaid to produce a false color map (d).

### *3.4.3 Score plots*

Sometimes the EDA tool itself can be enhanced to augment the information content presented. Such is the case with two dimensional score plots common to PCA analysis. This plot is often used for pattern recognition or classification purposes. When twenty to one hundred samples are plotted it is often possible to identify clustering of data points with or without the aid of additional symbol information. However, when twenty or one hundred thousand points from a hyperspectral image are plotted the relative abundance and importance of clustering is easily lost: one point may represent a single sample point, while another may in fact be one thousand sample points with identical score values. Esbensen & Geladi (1989) alleviated this problem by counting the number of points with each score pair (a two dimensional histogram) and mapping a color scale to the relative abundances. Figure 8 and Figure 23 show examples of color mapping, where green indicates a very few sample points and red or white indicates a large number of points. This color mapping helps to identify clusters of sample points with high populations indicating a large number of pixels with similar spectral characteristics. It is equally important to view the sparsely scattered points representing outliers or pixels with unique spectral features.

## 3.5 Summary of chemometrics

Advances in computational hardware and the development and application of chemometrics were necessary to transform the field of NIR spectroscopy into a useful analytical tool. This same combination of hardware and software tools must be further exploited to analyze and extract the information contained in hyperspectral NIR images. The excessive amount of data in each hyperspectral image is both a burden and an advantage.

The set of spectra selected to use for calibration model building must be carefully filtered (cleansed of outliers), identified and masked to represent only the constituent of interest (local model), and spectrally compressed to reduce physical or instrumental contributions to spectral variation (mean ROI spectra).

Processing results can be presented in new ways indicating population statistics and distributions (histograms) as well as incorporating spatial information (gray scale and false color mappings of score images, prediction images, or residual images). Two dimensional score plots can be color mapped to reveal relative abundance of pixels with similar or unique spectral characteristics.

# 4.0 Hyperspectral Image Transformation and Standardization

The signals acquired from a spectroscopic detector represent digitized values of voltage or current resulting from photons impinging on the surface of a photosensitive substrate. The magnitudes of these signals are wavelength dependent due to differences in quantum efficiencies of the detector, system optics, and illumination intensity variations. Standard Reference Materials (SRMs) are used to allow *transformation* of the instrument signals into reflectance or absorbance units, permitting comparisons between different sample spectra and spectral libraries. Multivariate techniques such as PLS can be used to create an additional layer of *calibration* models to permit the conversion of reflectance or absorbance spectra into concentrations of chemical constituents.

Both the reflectance transform functions and calibration models for prediction of chemical constituents remain valid only as long as the sample, instrument, and surrounding environment remain stable. Physical changes in the sample such as particle size or texture, changes in the instrument response function due to component aging or power fluctuations, or even changes in ambient room temperature or humidity may lead to erroneous results. Periodic instrument *standardization* is needed to ensure that spectroscopic data is consistent on a day to day basis from one instrument, as well as between instruments.

Instrument signal transformation, standardization, and calibration are common topics in spectroscopy. These processes can be applied to hyperspectral image spectra as well. This section of the thesis explores transformation and standardization issues that are unique to hyperspectral images. First, since a hyperspectral image contains thousands of spectra, how can the raw instrument signal be transformed to best ensure that a uniform response has been achieved for all pixel locations at all wavelengths within a single hyperspectral image? Second, can subtle image to image or day to day variations in imaging system response be detected in such an extremely large dataset, and if so, how can they be best corrected? Answers to these questions are dependent on having valid calibration standards to use for routine measurements. What standard reference materials should be used for calibration of hyperspectral images? These topics are discussed below.

## 4.1 NIR Standard reference materials

In the case of near infrared spectroscopy, spectral resolution varies considerably ranging from high resolution FT-IR (4 $cm^{-1}$ ~ 0.4nm) to very coarse resolution (10 to 20 nm) found in dispersive instrument configurations. The NIR region extends from 700 to 2500 nm (14300 to 4000 $cm^{-1}$) and the instrumentation is operated in both transmission and diffuse reflectance mode. Finding SRMs for this broad range of conditions is not easy. SRMs are needed to calibrate and correct raw spectra for variations in both wavelength and intensity axes. Spectral band

features in SRMs should not change because of the variation in spectral resolution of the instrumentation. Standards should be stable, inert, and insensitive to environmental changes such as temperature and humidity. They should also be easy to handle, easily cleaned, and be affordable.

### *4.1.1 Wavelength standards*

Several types of materials have been considered as NIR wavelength or wavenumber standards. Water peaks are abundant in the NIR region, but are typically unresolved multiplets. Choquette *et al.,* (2002) describe the history and release of the first certified SRMs. SRM 2517 and SRM 2519 are gas standards that contain many lines which can be fully resolved by high resolution instruments only. Polystyrene is inexpensive and contains highly reproducible but asymmetric bands, primarily occurring above 1600 nm (Choquette, 2000). Locating the peak center in all these materials is highly dependent on spectral resolution.

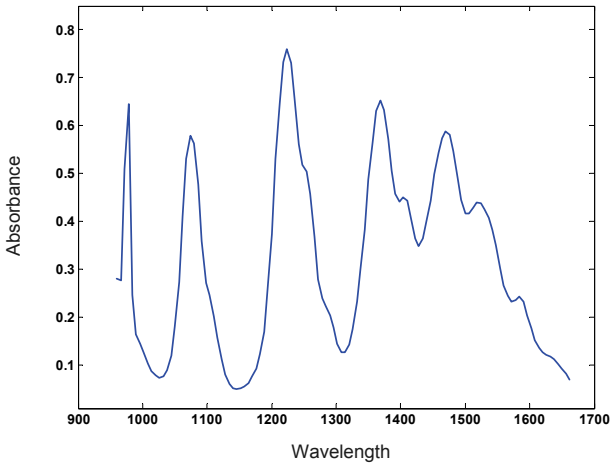


Figure 14 Spectrum of the prototype SRM 2035 standard reference material containing rare earth oxides.

The NIR optical properties of glass containing rare earth oxides were presented by Fuxi (1992). This led to a glass based SRM released by the National Institute of Standards (NIST), SRM 1920a, containing a mixture of erbium oxide ($Er_2O_3$), dysprosium oxide ($Dy_2O_3$), and holmium oxide ($Ho_2O_3$). Further development led to SRM 2035 and by adding a backing of sintered polytetrafluoroethylene (PTFE), SRM 2036 (Choquette *et al*., 2005). This material contains the rare earth oxides samarium ($Sm_2O_3$), ytterbium ($Yb_2O_3$), and holmium ($Ho_2O_3$). Choquette, O'Neal, & Duewer (2003) examined the stability of these glass materials as a function of temperature. These SRMs have also been used to explore comparisons of spectrometers and peak fitting or spectral preprocessing effects on peak center locations (Isaksson *et al*., 2003). Specially cut pieces of a prototype glass for SRM 2035 were obtained from Steve Choquette, NIST for use in this thesis. A

typical spectrum for this material is displayed in Figure 14. In this case the spectrum is the mean of over 12,000 spectra contained within a sample ROI of a diffuse reflectance hyperspectral image of the glass material positioned on top of 99% reflectance material.

### *4.1.2 Reflectance standards*

Spectroscopic reflectance standards are generally provided in one of three forms: powdered, thin films or coatings, or solid tiles. Color ceramic tiles are popular SRMs for calibration of visible light (CERAM, Staffordshire, UK). Similar white ceramic tiles can also be used in NIR applications; however reference materials made from MgO, $BaSO_4$, or PTFE are more common. Both $BaSO_4$ and PTFE materials are used in powdered from as pressed disks, or as surface coatings for integrating spheres (Weidner & Hsia, 1981). The variability of test spectra due to powder batch, instrumentation, and operator differences was reported by Barnes & Hsia (1995).

For imaging applications it is desirable to have material that exhibits high Lambertian reflection, diffuse reflection that is uniform in all angles. The predominant material in use is Spectralon, a sintered PTFE material (Labsphere, USA). This material is also available with increasing amounts of black carbon added to produce gray reflectance standards. Spectralon tiles with 99, 75, 50, 25, and 2 percent reflectance values were used extensively in this thesis. These gray body reflectors are appropriate for area integrating spectrometers; however, at some level of magnification this material no longer appears homogenous. This can cause problems for image calibration. For this thesis, this issue was partially alleviated by randomly moving the Spectralon tiles during image acquisition.

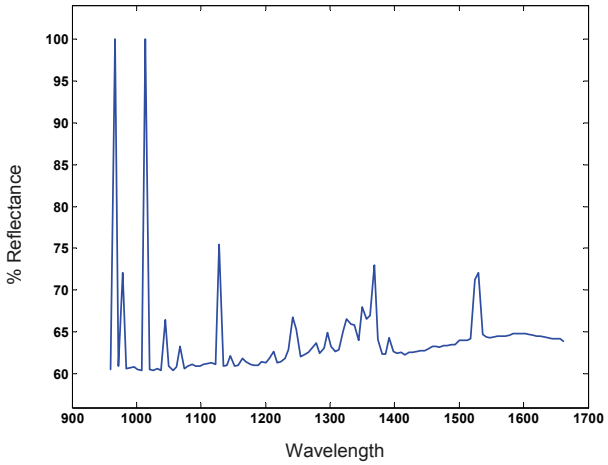


Figure 15 Emission spectrum of a fluorescent light tube averaged from 5000 sample spectra, scaled as percent reflectance.

### *4.1.3 Additional Considerations*

One source of spectral measurement error is the detection of stray light within the spectrometer system. This is especially important in imaging applications where high specular reflections from regions outside the intended sample field of view can enter the camera. It is important, therefore, to position image samples on a background with minimal reflection. Silicon carbide sandpaper provides an ideal sample background for this purpose; it has a very low reflectance throughout the NIR region, the surface texture helps to prevent sample movement, and it is inexpensive and disposable.

A possible inexpensive radiation source for NIR wavelength calibration is the light emitted from 'low energy consumption' fluorescent lamps readily available for home use. This light can be measured either directly or reflected from a Spectralon surface. Figure 15 displays the average spectrum measured from 5000 spectra in a hyperspectral image region of a 20 W fluorescent lamp. The peaks are very sharp; however, the effects of temperature and power supply fluctuations on peak position are not known.

New SRMs are needed specifically for hyperspectral imaging which appear homogenous at different scales of magnification. It is not necessary that these materials have flat response spectra, as with Spectralon. This is actually a disadvantage since very little wavelength information is conveyed. It is only necessary that a set of standards span the full spectral space of the samples to be measured. New standards that span the sample space and reflectance space need to be further investigated. Standards for spatial calibration should be considered as well; spherical aberration, vignetting, and depth of field are important issues that relate to spatial resolution, but are dependent on wavelength as well.

## 4.2 External standards

Paper I describes in detail the use of a set of five Spectralon tiles to build models for transforming raw instrument signals into reflectance. These models were based on acquisition of complete hyperspectral images of 99, 75, 50, 50, and 2 percent reflectance tiles. The objective was to transform the raw signal from any future sample images into reflectance. The average spectrum from each hyperspectral image can be used to create transform models for each wavelength channel, accounting for any wavelength response dependencies in the camera. These models are termed *global* models. To account for spatial variation in illumination and other system responses, independent reflectance transform models were also computed for every pixel location at every wavelength channel, termed *local* models.

These two broad classes of transformation models were further sub-divided based on the type of model computed. Calibration reference spectra were available for each of the Spectralon SRMs, consequently regression models were computed where the independent variables **x** represent the five hyperspectral image spectral values, and the dependent **y** variables were computed from

polynomial interpolation of the supplied reference spectra. Five sample points were available, so that both *linear* and second order (*poly*) regression models could be computed. Since baseline offsets are determined automatically by regression, no dark current or background measurement is needed. However, for comparison purposes, dark current images were also acquired and used with the 99 percent reflectance image to create single point transforms, termed *simple* models.

Figure 16 (Paper I) summarizes the general results of these approaches. The second order model permits the best approximation of the target reference spectra (a) and the local models (b) reduce the noise contributions sustained at each pixel.

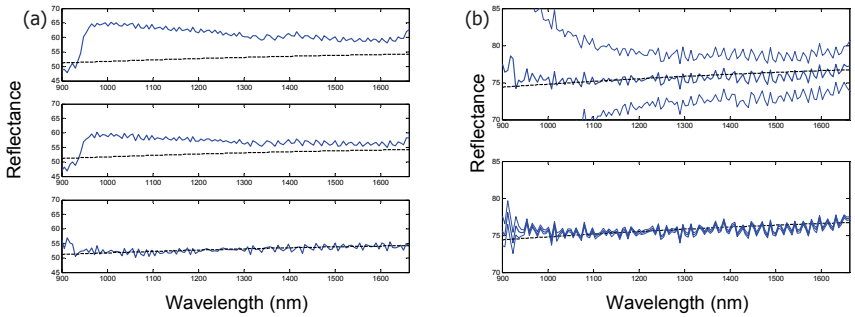


Figure 16 (a) The average spectra for 50 percent reflectance Spectralon based on simple (top), linear (middle), and second order (bottom) regression models. (b) The average spectra with standard deviation bands for 75 percent Spectralon based on second order global models (top) and individual pixel models (bottom). The targeted reference spectra are indicated by the dashed lines.

## 4.3 Internal standards

To correct for these instability issues, the instrument standardization techniques Direct Standardization (DS) and Piecewise Direct Standardization (PDS) were introduced (Wang, Veltkamp, & Kowalski, 1991; Shenk & Westerhaus, 1991). Here a set of carefully selected calibration standards are measured on two different instruments termed a *slave* and a *master*. The two resulting sets of spectra are used to compute a transfer function that allows the conversion of spectra from the slave instrument into approximations of the equivalent spectra, had they been acquired on the standardized master instrument. Computation of the transfer function is based on the spectra from the calibration standards; however, once computed it can be applied to transform any subsequent sample spectra. While this process does improve spectral stability, it is time consuming and there is uncertainty about when or how often to perform a new standardization.

The use of SRMs embedded within the image field of view was investigated for instrument standardization. The hyperspectral images acquired for use in this thesis work contained over 80,000 pixels or spectra; some of the spectra could be used for calibration purposes, while retaining large areas for sample imaging. Initially four tiles of Spectralon were positioned in the corners of the field of view,

however the metal frames enclosing the tiles used image space and their specular reflection may have contributed to stray light problems. A gray scale was obtained that contained eight pieces of Spectralon laminated together that could be positioned along one edge of the image (TestElek Svenska AB). This calibration scale ranged from 99 to 2 percent reflectance, but its 10 cm length prohibited the viewing of all eight segments at one time. Consequently, some of the material was excluded from the field of view. Figure 17 shows a typical image indicating the placement of this scale.

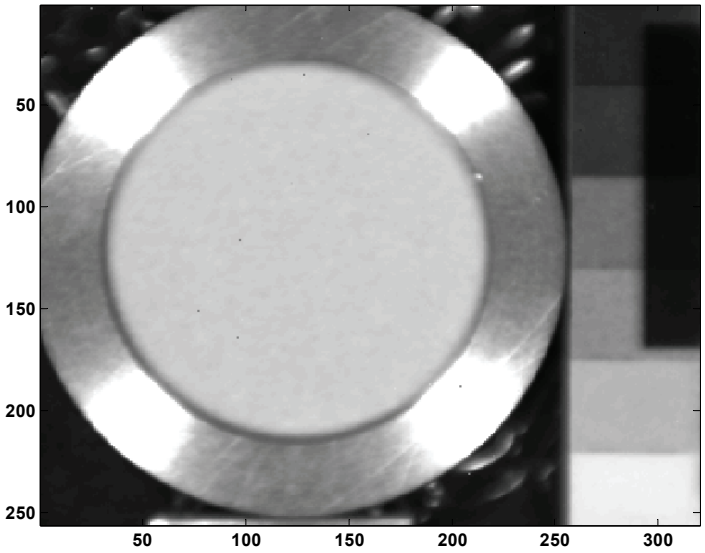


Figure 17 A typical image (320 x 256 pixels) indicating the placement of internal standards for intensity calibration (gray scale on right) and wavelength calibration (dark rectangle on top of gray scale). The sample is contained in a circular holder with glass window. The axes labels indicate pixel location.

Since the gray scale was included in all images, instrument standardization based on signal intensity could be achieved. But the Spectralon materials purposely have a very flat spectral response and are therefore inappropriate for wavelength calibration. A rectangular piece of the prototype SRM-2035 glass was obtained from NIST which could be positioned directly on top of the Spectralon gray scale. This configuration provided a series of wavelength calibration spectra with decreasing signal intensity. This material can also be seen as the dark rectangle on the far right in Figure 17.

Corrections based on these internal standards are discussed in both Papers I and II. In general it was found that the intensity corrections had very little effect on quantitative results. However, the spectra were useful in monitoring the health of the camera system. The median spectra from a region of the 99 percent standard for a series of hyperspectral images acquired in one day were analyzed with PCA. The resulting score plot is displayed in Figure 18. The two clusters of sample

points can be attributed to image acquisition times occurring either before or after an automatic upload of the filter settings in the camera. This change in the filter had minimal apparent effect on the prediction results from sample spectra, but was readily apparent in the diagnostic plot. Unfortunately the glass standard was not available at the time that these images were made, however similar PCA plots of spectra from the glass standard from images acquired on later days were also periodically examined. In those cases, variation in score values appeared to be random, with no time dependencies indicated. This approach utilized the spectral intensities at all wavelengths. The location of peak centers could be compared as well but was not attempted.

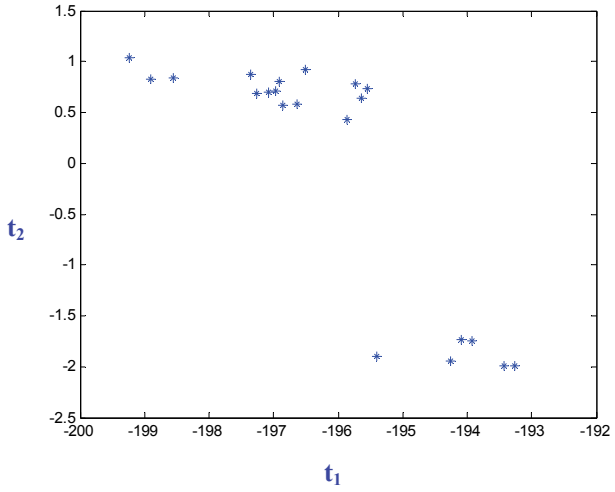


Figure 18 Score plot of 99 percent reflectance internal standard from 21 hyperspectral images. The two clusters of points correlate to image acquisition times before and after an update to the camera filter system.

# 5.0 Outlier Detection

An *outlier* is an observation that lies at an unusually large distance from other values observed in a small random population sample. Outliers can distort simple population statistics such as the mean and standard deviation, and have high leverage effects on chemometric techniques (Martens & Næs, 1989; Beebe, Pell, & Seasholtz, 1998; Næs *et al.*, 2002). This was demonstrated by the effect on linear regression indicated in Figure 10 where a single point greatly perturbed the slope of the regression line. It is important to identify dataset outliers and selectively remove them from both calibration model generation and prediction processes. Since hyperspectral images contain millions of data values, it is also important to utilize automatic outlier detection and removal schemes. Several different approaches investigated in this thesis are described below. To correctly determine any abnormal outliers, what is normal must first be characterized. This characterization includes identifying the expected range in data values, an estimate of signal to noise ratios, and especially with multivariate data, the shape and orientation of sample distributions.

In the case of the hyperspectral NIR imaging system used in this thesis, the Spectralon standard reference materials used to create reflectance transform models could be used to assess the signal and noise structure of the system as well. The series of five large Spectralon tiles were first imaged and used to create a second order reflectance transform model for each wavelength at every pixel. This model was then used to transform the five acquired images into reflectance values. Any spatial variations due to instrumentation should have been removed by this process, resulting in hyperspectral images with very flat or uniform wavelength channels. The average spectrum from each image should agree with the reference spectrum provided with the standard reference material.

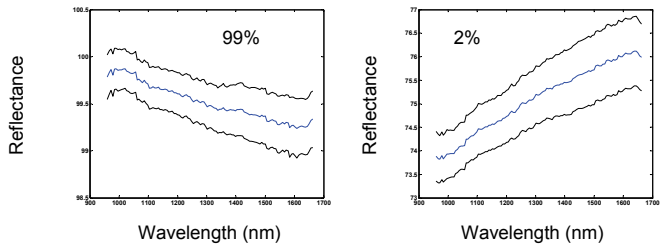


Figure 19 Mean spectrum with standard deviation computed from the entire hyperspectral image, 99% reflectance standard left, and 2% right.

Figure 19 displays the mean spectrum and standard deviation for both the 99 and 2 percent SRMs. Both plots have the same vertical scale of 2 percent reflectance. The standard deviations for both materials are fairly consistent throughout the full spectral range with a gradual increase from 0.2 to 0.3 percent

increasing with longer wavelengths. These two materials span the extremes of nearly total reflectance to total absorbance and suggest a rather homoscedastic noise structure.

The 81,000 points within the 1200 nm wavelength channel are displayed as histograms for each of the five Spectralon standards in Figure 20. Again the 99 and 2 percent SRMs appear quite consistent, having distributions with similar shapes and sizes. However, the intermediate reflectance SRMs have a substantial wider distribution. For comparison purposes, all five histograms have the same vertical scale and a 10 percent wide horizontal scale.

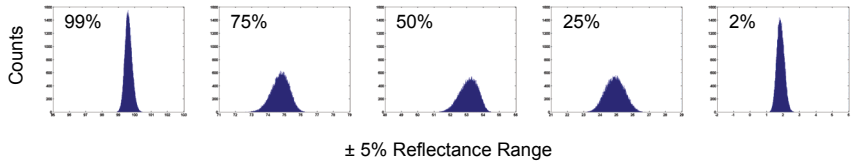


Figure 20 Histograms of 81,000 points measured at 1200 nm in each of the Spectralon standards indicates a substantial increase in noise with the blended materials used for 25, 50, and 75 percent reflectance.

What caused this increase in sample variance? The two extreme reflectance standards, 99 and 2 percent, are manufactured from very homogenous PTFE derived materials; the intermediate valued reflectance standards are blends of these two materials. With course magnification or sample spinning often found in integrating spectrometers, these two materials are averaged together. The hyperspectral images used for Figure 19 and Figure 20 were made from fixed or stationary samples where the spatial resolution of the image captured the inhomogeneity of the blended materials. Based on these conclusions, the Spectralon materials were either spun on a bearing (Volvo, Sweden) or randomly moved by hand during image acquisition. This physical averaging effect generally decreased the peak width of the signal distribution, but introduced other complications such as doublet formations, possibly due to vibrations or tilting of the sample surface during sample movement. Homogenous materials or better solutions to random sample movement need to be developed.

## 5.1 Hardware bad pixels

Detectors such as the InGaAs detector used in this thesis are notorious for having random elements which respond with erroneous data values. These elements in images are known as *dead* or *bad* pixels. Often these points occur with either a constant zero or maximum signal value, but elements that are simply stuck at an intermediate value may occur as well. The detector in the hyperspectral imaging system for this thesis used a 12 bit analog to digital converter, with data values ranging between 0 and 4095. This instrument was tuned with a 99 percent

reflectance tile so that the maximum expected signal counts were typically 3900, thereby avoiding signal saturation. The measured dark current or background signal from this detector had a typical value of 300 counts. Consequently a simple thresholding approach was used to identify any pixels with an unusual value; any spectrum containing a data value less than 100 or greater than 4000 was removed from further processing. This cleaning process removed the problematic data outliers with extreme signals, but it did not detect or remove the pixels with constant intermediate values within the acceptable data range.

## 5.2 Distance to mean spectrum

One of the simplest outlier tests was based on the Euclidean distances between spectral data sample points. Data subsets were first selected utilizing spatial masks identifying regions of interest (ROIs) within sample images. These rectangular or circular regions were selected such that each individual ROI contained only a single class of sample data. When the selected spectra contained within such an ROI mask are viewed in a K dimensional feature space (K equals the number of variables or wavelengths), their distribution should appear as a cluster or ball of data points. One test for outliers is based on the Euclidean distance to the center of this cluster. This test is depicted in Figure 21 for two dimensions.

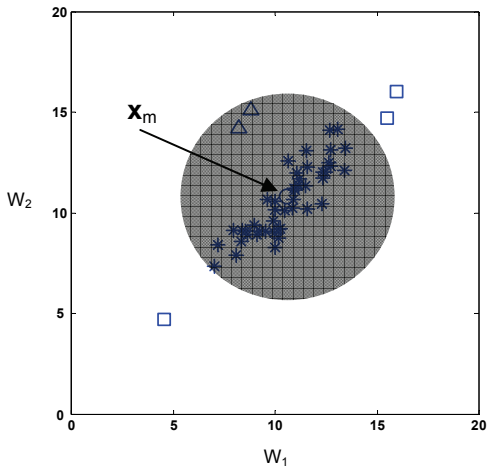


Figure 21 Outliers can be identified based on their Euclidean distance to the mean sample point $\mathbf{x}_m$ indicated by the small circle symbol. The gray region indicates a hypersphere boundary with a radius of three standard deviations based on the distances of all sample points. The sample points marked with squares are outside this boundary and are therefore identified as outliers. The samples marked with triangles may also be outliers, but fall within the hypersphere boundary.

The center of the cluster, $\mathbf{x}_m$, indicated by the circle symbol was first computed as the mean of all spectra contained in the sample subset $\mathbf{X}_c$. The distance d for a single sample spectrum $\mathbf{x}$ to this center point was then computed:

$$d = ((\mathbf{x} - \mathbf{x}_m)^T (\mathbf{x} - \mathbf{x}_m))^{1/2} \quad [15]$$

An initial outlier test was based on the computed distances **d** for all spectra contained in $\mathbf{X}_c$. An estimate of the mean $m_d$ and standard deviation $s_d$ of the population of distances **d** was computed and any spectra with distances greater than three standard deviations from the mean distance value were removed from $\mathbf{X}_c$. However, the location of the mean spectrum $\mathbf{x}_m$ is heavily influenced by outliers, which led to an improved detection and cleaning process: after removing the detected outliers a new $\mathbf{x}_m$, **d**, $m_d$, and $s_d$ were computed from the cleaned $\mathbf{X}_c$. The new distances **d** were tested again for additional outliers based on the new population statistic estimates. As additional outliers are removed, the mean $m_d$ may shift and the standard deviation $s_d$ will shrink. Consequently this detection and cleaning process was repeated iteratively until no more outliers were detected.

This simple Euclidian distance cleaning process gives equal weight to all wavelength variables and assumes equal variances as well. Outlier detection is effectively based on the identification of any outliers outside a hypersphere with a radius of three standard deviations. The three sample points marked with squares in Figure 21 were identified as outliers. The sample points with the triangle symbols could also possibly be considered outliers, but their distances were within the limits of the hypersphere.

## 5.3 Angle to mean spectrum

If a sample set contains many spectra with a common chemical constituent, any variations in the concentrations of this constituent will be observed as an elongated clustering of feature space points. The elongated axis will be aligned with a vector representing the spectral values of this constituent. Such a data structure is exemplified in Figure 21 and Figure 22. This variance in concentration contributes to a larger sample variance which increases the radius of the hypersphere boundary used in the outlier distance test. A second outlier detection test was inspired by the collinearity of such expected components.

The similarity of two spectra can be quantified as the cosine of the angle between their vectors in feature space. This value can be used to test for outliers. As in the distance based outlier test, a mean spectrum $\mathbf{x}_m$ was first computed from all sample spectra $\mathbf{X}_c$. The cosine of the angle α between any spectrum **x** and $\mathbf{x}_m$ can then be computed:

$$\alpha = \mathbf{x}^T \mathbf{x}_m \, ( \|\mathbf{x}\| \, \|\mathbf{x}_m\| )^{-1} \quad [16]$$

Here the spectra are normalized to unit length as indicated by the double bar operator.

The mean value $m_\alpha$ and standard deviation estimate $s_\alpha$ were then computed from the entire collection of angle cosines **α**. Outlier detection and trimming of samples with an angle cosine greater than three standard deviations from the mean was

performed. As with the distance detection approach, this cleaning process was repeated iteratively, refining $\mathbf{X}_c$, $\mathbf{x}_m$, $m_\alpha$, and $s_\alpha$, until no further outliers were detected. Figure 22 represents this angle based outlier detection. In this case the dissimilar points indicated by the triangle symbols are detected as outliers, while the sample points with square symbols previously identified as distance based outliers are now deemed acceptable.

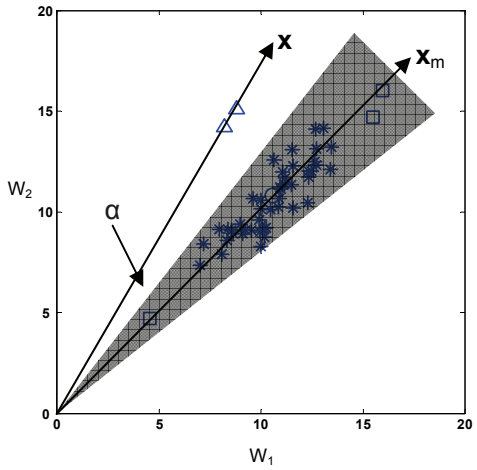


Figure 22 Outlier detection can be based on spectral similarity as indicated by the angle α between the vectors passing through sample point **x** and the mean spectrum $\mathbf{x}_m$. The gray region represents a three standard deviation wide region of acceptable points, based on the angle cosines of all samples points. The samples with triangular symbols are now identified as outliers, while the samples represented with square symbols are acceptable points.

## 5.4 Prediction extremes

The outlier detection tests of faulty hardware, or distance and angles in feature space, are all descriptive of spectral information obtained from wavelength data, or independent variables. Outliers may also be observed as unacceptable values in predicted values, or dependent variables. When validating a PLS calibration model, both calibration and test set prediction values and residuals should be examined. As with the spectral distance and angle tests, estimates of the population means and standard deviations may be computed to set limits for outlier detection.

## 5.5 Example

The following example demonstrates the advantages of the outlier detection and cleaning process. A hyperspectral NIR image was made of four Swedish cheeses surrounded by four internal standard Spectralon reference tiles. All spectra were transformed to reflectance based on a second order transform model. Figure 23 indicates the effect of the sequential application of masks and the PCA score plots of the spectra remaining in the image dataset.

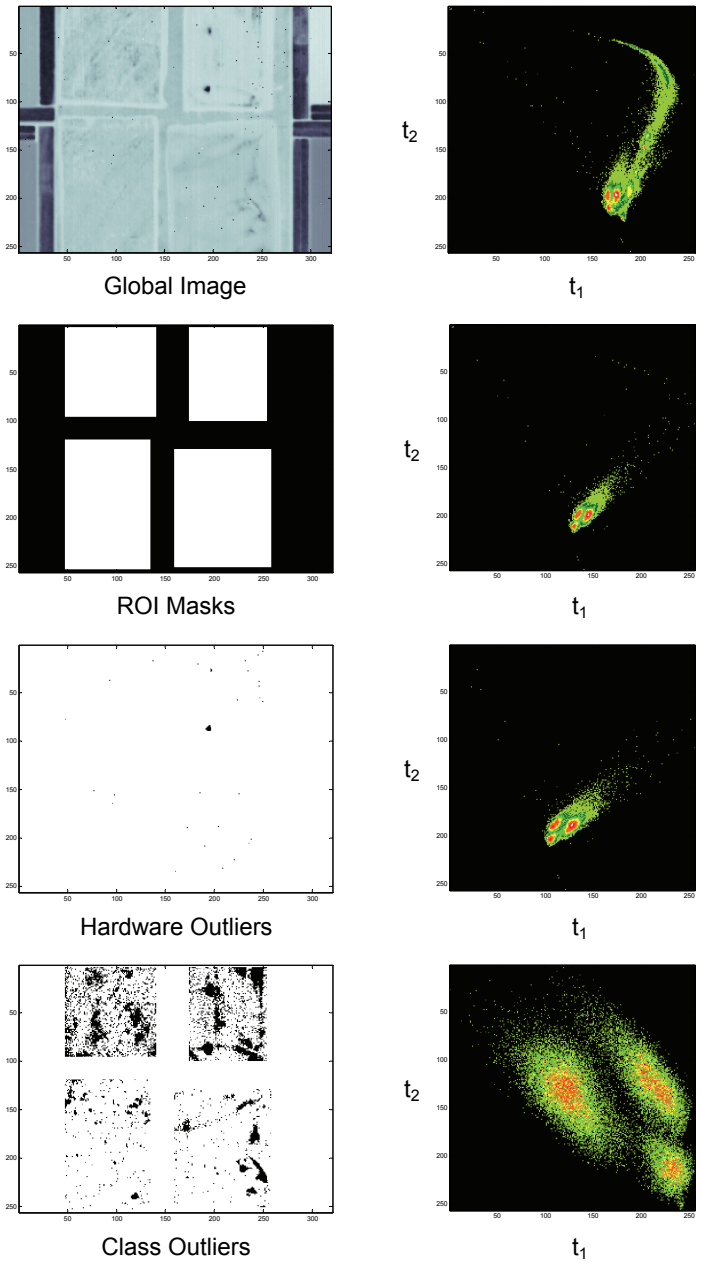


Figure 23 The sequential application of masking and outlier detection and removal is applied to the hyperspectral image of four cheeses. A representative wavelength channel image and score plot of all 81920 spectra is presented in the top row. A region of interest mask in the second row reduces the sample pixel count to 40,145. Thirty two hardware outliers are identified and removed in the third row. 6461 outliers were detected and removed in the last row, based on a class distance cleaning of each of the four cheese regions. The end result is a score plot with three well resolved classes. Two of the cheeses were severely overlapped and could not be resolved.

The top row displays an image of one wavelength channel showing the location of the four cheese samples. The long streaming tail formation evident in the color intensity mapped score plot of all 81920 spectra is due to the internal standard reference materials and their frames.

The four white rectangles in the second row indicate region of interest mask areas to select for the four cheeses. The spectra associated with the black pixels were removed from the data set leaving 40,145 sample spectra. This ROI masking removes the streaming tail structure from the score plot, but many scattered outlier points are still evident.

Hardware thresholding was applied to the remaining spectra, resulting in 32 randomly located spectra identified with data values out of range. These pixels are indicated in the mask on the third row. Even after removal of the faulty spectra from the data set, the score plot still indicates many remaining outliers.

Spectral class distances were computed for the remaining pixels in each of the four cheese classes. Each class of spectra was processed using the iterative distance detection algorithm based on class population mean and standard deviation. In total, 6461 outliers were detected and removed, indicated by the black pixels in the mask on row four. This may sound like a significant number of samples to throw out, but it is less than 8 % of the original image. However before discarding, it is important to examine any outliers for causes of data faults. In this case these pixels had very high values, probably due to specular reflection. Removal of this last set of class outliers resulted in a major improvement in the score plot. Three very well defined clusters of points can now be seen which correlate with the cheese types. Two cheese types were highly overlapped and could not be resolved as separate classes.

The example given here was analyzed before the development of the class based angle similarity detection algorithm. PLS-I models were made for the determination of fat and protein. Iterative statistical analysis of the predictions of the calibration data set resulted in identification and removal of 41 additional outlier points. This combination of localized ROI masked data and detection and removal of dataset outliers assured generation of the best possible PLS calibration model.

## 5.6 Summary of outlier detection

No clear approach is adequate for detection and removal of all outlier data; each of the spectral based processing approaches, hardware thresholding, class distance, or class similarity, was insufficient by itself. The combination of all three approaches did however effectively screen most outliers. These processes could be totally automated; however it is important to manually examine the outliers that are being removed, as well as the clean data, to make sure no outliers remain. Additional spectral preprocessing further complicates things. It was found that applying treatments such as derivative transformations to previously cleaned datasets, often resulted in datasets with new outlier types. Because of this, the outlier detection process was repeated after preprocessing treatments as well.

It should be cautioned that the estimates of population statistics such as mean and standard deviation were often made on distributions that were clearly not normal distributions. These statistics were still useful for characterizing distribution differences; however, their use should not be confused with normal distribution statistics.

Alternative outlier tests could be applied as well. Computation of PCA and Mahalanobis distances could be performed to first suppress unwanted noise. The PCA loading vectors are also susceptible to leverage effects by outliers, so an iterative approach should be utilized here too. This could be quite computationally exhaustive with large hyperspectral data sets. An iterative Singular Value Decomposition (SVD) approach could be considered to speed up this approach and warrants further research for more complex spectral sets containing multiple classes.

Regardless of the detection scheme used, the sheer number of spectra available in hyperspectral images provides two distinct advantages for outlier detection. First, since sample population size is quite large, population statistic estimates are reliable. Second, this large sample size permits the liberal removal of outliers; if hundreds or even thousands of samples are thrown out, tens of thousands of samples still remain. This permits the selection of very clean calibration sets which lead to more reliable and robust calibration models.

# 6.0 Spectral Preprocessing

Individual spectral outliers are data vectors containing errors or dissimilar data values that cause them to appear abnormal or substantially different from the normal or expected population of spectra. Less severe sources of data variance may also occur which may effect many or all spectra in the population contributing to an overall shift or disturbance to the quality of the dataset. Sources of variance could be instrumental noise, sample preparation issues, or simply sample contamination or introduction of a new chemical component. Various filters or transformations may be applied to these spectra to reduce the effect of these variances. Transformations may be applied to reduce or counter the effect of noise, account for baseline offsets or slopes, or simply change non-linear components of the signal. Collectively these data transformations are called spectral preprocessing treatments.

While these treatments were initially developed for processing small sets of individual spectra from integrating spectrometers, they can also be applied to the thousands of spectra within hyperspectral images. Two related questions were considered in this thesis: First, how do preprocessing treatments of the spectra affect prediction results unique to hyperspectral images? Second, what additional sample information can be gained from the combination of preprocessing treatments with the spatial information provided by hyperspectral images? The first question is addressed further in papers II and III, while the second is explored in paper IV. Results are summarized below. For general discussions of spectral preprocessing treatments, see Katsumoto *et al*. (2001), Næs *et al*. (2002), and Thennadil & Martin (2005).

## 6.1 Linearization transforms

The transformation process described in section 4.2 of this thesis was used to convert instrument signal counts to reflectance units; however, the resulting reflectance values are not directly proportional to chemical constituent concentrations, frequently the objective of spectral analysis. Two spectral pretreatments are often applied to linearize reflectance responses: conversion of spectral units to absorbance, or application of the Kubelka-Munk transform.

### *6.1.1 Absorbance*

The Beer-Lambert law (Eq. [7] section 3.3.1) expresses the linear relationship between spectral absorbance and concentration. For a given analyte at a particular wavelength the absorbance A is defined as:

$$A = - \log_{10} ( I_T / I_0 ) = - \log_{10} ( T ) \qquad [17]$$

Where:
$I_0$ is the intensity of incident light
$I_T$ is the intensity of light transmitted.
$T = I_T / I_0$ is known as the transmittance $( 0 \leq T \leq 1 )$

For reflection spectroscopy a similar expression is used:

$$A = - \log_{10} ( R ) \quad \text{or} \quad A = 2 - \log_{10} ( R_P ) \qquad [18]$$

Where:
$R = I_T / I_0$ is known as the reflectance ( $0 \leq R \leq 1$ )
Or in alternate units, $R_P$ is percent reflectance ( $0 \leq R_P \leq 100$ )

This equation was used to transform the reflectance values to absorbance at each wavelength of the hyperspectral image spectra. This linearization of reflectance values is only valid for low analyte concentrations, i.e. high reflectance values.

### *6.1.2 Kubelka-Munk*

The Kubelka-Munk theory was developed to account for the effect of both absorbance and scattering in thin films of paint. (Kubelka & Munk, 1931). Equation 19 shows the relationship between the absorbance and scattering constants K and S and the diffuse reflectance R from an infinitely thick material:

$$K/S = (1-R)^2 / 2R \qquad [19]$$

As with the absorbance transformation, this equation was applied directly to each wavelength of the hyperspectral image reflectance spectra.

The K value is primarily a function of the chemical composition, while S is dependent on physical and non-homogeneity properties of the sample. When R is less than 60% (or $K/S > 0.13$), equation 19 deviates from experimental results (Kortum, 1969). This suggests that the Kubelka-Munk transform should be applied when scattering dominates over absorbance. This can be achieved by grinding samples to very fine powder, increasing scattering. Samples can also be diluted with non-absorbing materials such as potassium bromide. An alternative linearization treatment has been proposed by Švedas (2004) but was not considered in this thesis.

## 6.2 Additive transforms - derivatives

Changing spectral baseline offsets and slopes are considered as additive processes. For example changes in instrumentation such as lamp intensity, temperature, or detector response, or changes in sample orientation, particle size distributions, or packing may result in a background signal that is added equally throughout the spectrum, or possibly uniformly as a function of wavelength. These constant additive effects can be compensated for by applying first and second derivative transforms (Giesbrecht, McClure, & Hamid, 1981; Norris & Williams, 1984; Hopkins, 2001). Derivative spectra can be computed based on the differences of adjacent data points; however, this can greatly magnify noise contributions. When discrete spectral data points are evenly spaced, derivatives can be computed using standard Savitsky-Golay polynomial filters (Savitzky & Golay, 1964; Steiner,

Termonia, & Deltour, 1972; Madden, 1978). These filters have an additional smoothing effect resulting from fitting the multiple data points used within the filter window. The use of digital derivatives of higher orders can also be used to assist with the determination of peak widths and intensities (Morrey, 1968). Such information may be useful for matching of spectra to spectral libraries.

Derivative transforms can be applied to both reflectance and absorbance spectra. In this thesis derivatives based on absorbance spectra that retained their linear relationship with constituent concentration were used for prediction analysis. The analysis of scatter effects was based on the transformation of reflectance spectra. The hyperspectral images considered here typically had 118 wavelength channels. Second order five point polynomial filters were used for both first and second derivative transforms. Window sizes were determined empirically based on the visual examination of sample plots of transformed spectra. These values retained the maximum spectral resolution while minimizing noise enhancement effects.

## 6.3 Multiplication transforms - scatter corrections

Martens, Jensen, & Geladi (1983) and Geladi, McDougall, & Martens (1985) proposed an alternative correction for scatter effects, Multiplicative Scatter Correction, (MSC). A target spectrum is first computed, typically the mean spectrum $\mathbf{x}_m$ from the sample data set. For each sample spectrum $\mathbf{x}$, it is assumed that the scatter effect is constant for all wavelengths. A plot of $\mathbf{x}$ vs. $\mathbf{x}_m$ for all wavelengths should therefore result in a straight line. The offset $b_0$ and slope $b_1$ estimates from a linear regression of $\mathbf{x}$ vs. $\mathbf{x}_m$ can be used to compute a corrected spectrum, $\mathbf{x}_c$:

$$\mathbf{x}_c = (\mathbf{x} - b_0) / b_1 \qquad [20]$$

This correction process computes a different offset and slope unique to each sample spectrum. But the scatter effect may not be exactly the same for all wavelength ranges. Isaksson & Kowalski (1993) proposed correcting the spectral value at each wavelength with independent offset and slope correction terms. These terms are also computed by linear regression, using a small window of data centered about each wavelength. This technique is called Piecewise Multiplicative Scatter Correction (PMSC). Another variant, Extended Multiplicative Signal Correction (EMSC) was presented by Martens & Stark (1991) and Martens, Nielsen, & Engelsen (2003). Results from testing both the MSC and PMSC techniques on hyperspectral image spectra are presented in this thesis. These results are detailed in Paper IV, and summarized in section 8.2.

## 6.4 Other transforms

The Standard Normal Variate (SNV) transform was proposed by Barnes, Dhanoa, & Lister (1989) to correct for multiplicative scatter effects. Although this technique does not require the use of a target spectrum, it has been shown to be linked directly to the MSC transform (Dhanoa *et al.*, 1994). This transform normalizes each spectrum $\mathbf{x}$ by first subtracting its mean value $x_m$ followed by

dividing by the standard deviation $x_s$. This process effectively creates a common baseline offset (zero) and variance (one) for all spectra. Two additional normalization techniques were considered: the 1-Norm and 2-Norm. These normalization transforms scale each spectrum **x** such that the area of **x** is one (1-Norm) or the length of **x** is one (2-Norm). These three normalization techniques were explored in this thesis.

New signal correction techniques have been more recently introduced for reducing the noise contributions that are not correlated to dependent or response data. Orthogonal Signal Correction (OSC) was first introduced by Wold *et al*. (1998) with subsequent improvements suggested (Westerhuis, de Jong, & Smilde, 2001; Trygg & Wold, 2002; Trygg, 2003). A comparative analysis application was presented by Pizarro *et al*. (2003). In addition to improving spectral quality by the application of spectral pretreatments, multivariate analysis techniques have been extended to make them more robust to outliers and noise. This work has been described in summary by Liang & Kvalheim (1996) and Møller, von Frese, & Bro (2005).

## 6.5 Applied testing

Three different collections of samples were investigated to explore the effects of preprocessing treatments and multivariate analysis of hyperspectral NIR image spectra:

1. A designed experiment sample set consisting of standard mixtures of sugar, citric acid, and salicylic acid was carefully prepared using gravimetric techniques. This set of Tri-Mix samples was created to provide mixture samples with accurate response variable data, the concentrations of the individual constituent ingredients. The ingredients chosen had a combination of unique and highly overlapping spectral features.

2. A set of commercially available Swedish cheeses were obtained and imaged to use for testing the effect of preprocessing treatments on the predictions of fat, protein, and carbohydrate content. This sample dataset represented a valuable application of NIR imaging to food analysis. The response variable values were obtained from product packaging labels, a much less accurate source than the artificially blended gravimetric sample set. A parallel set of more accurate fat and protein reference values was obtained for some of the cheeses using standardized analysis protocols.

3. To investigate relationships between light scattering and preprocessing treatments a set of samples containing a range in particle size fractions of commercial grade salt and sugar were imaged. Eight sieves with mesh sizes ranging from 400 to 63 micrometers were used to sort the particles into nine size fractions.

Replicate images were acquired of the different sample sets to facilitate splitting of data into calibration and test sets. For comparison purposes all three sample set types were also measured with a grating based integrating NIR spectrometer. Results of concentration predictions from PLS regression models of the Tri-Mix samples and the cheese samples were presented in Papers II and III. Variations in light scattering due to particle size differences were examined in Paper IV. Additional information is provided in section 6 of this thesis, Applications.

# 7.0 Java Implementation of Multivariate Image Analysis - JIMIA

Supplementing spectroscopic information with detailed spatial or positioning data is the primary benefit of hyperspectral imaging. The multivariate statistics techniques described in this thesis enable classification and quantification of chemical constituents at every individual spatial unit in the image, summarized by both image maps and distribution profiles. Although these techniques are impressive in their utilization of all spectral channels, they are at the same time very constrained and limited in their approach; they make no use of the spatial information in either the modeling or prediction stages.

This approach is contrary to how the human mind processes visual information; when we see a piece of fruit in a basket we immediately classify it as either familiar or new, based on our *simultaneous* observation of *both* color and shape. Part of the research effort behind this thesis was an attempt to emulate this dual spectral – spatial processing in computer software.

## 7.1 Software design goals

### *7.1.1 Spatial analysis*

Many approaches exist for computerized analysis of two dimensional gray scale images. Spatial filters are available to sharpen, blur, or enhance image contrast. Computer vision topics include edge detection, feature recognition, texture analysis, and particle size counting; however, these techniques make minimal use of any spectral information; they simply operate on a single gray channel image.

### *7.1.2 Spectral analysis*

The multivariate statistics techniques described so far in this thesis utilize the spectral information, but ignore the spatial information. The order of the spectra contained in the unfolded hyperspectral image can be totally randomized, yet yield identical PCA or PLS loading vectors.

### *7.1.3 Hybrid approaches*

How can these two dissimilar approaches be combined? One solution is to first perform PCA to compress the wavelength channels, providing score images that contain maximum signal variance; spatial analysis can then be applied to the score images. A second solution is the inverse process; spatial based feature detection can be used to select spectra subsets followed by local PCA modeling. But is there a way to better enable *simultaneous* utilization of both spectral and spatial components? In 1989 Esbensen & Geladi introduced an interactive analysis technique for Multivariate Image Analysis (MIA) they termed ‘brushing’. Software was developed to present both the score images and score plots resulting from the application of PCA to a hyperspectral image. Pixels selected in a score image were mapped to their equivalent points in the score plots; alternatively

points in the score plots could be selected and their original spatial coordinates identified back in the score images. This solution begins with a purely spectral technique, PCA of all spectra, but then offers simultaneous viewing of spatial and spectral components.

This interactive approach to MIA was the starting design component of the JIMIA program. Additional functionality was incorporated incrementally to allow greater flexibility in the selection of data points used for highlighting or brushing, as well as enhanced methods for the selection of spectral subsets utilized for PCA modeling.

## 7.2 Software implementation

Several programming languages and environments were considered for this project. Java was selected as the core language for several reasons. First, because of limited resources, extensive function libraries were needed to provide some of the basic interactive and computational functionality. These libraries are described below. Second, although the program was developed in a Microsoft Windows platform, it was initially intended that the program would function cross-platform, and execute on both Apple and Linux systems. Third, it was important to use an object oriented language that would execute efficiently and support the very large memory requirements in processing hyperspectral images. And finally, since this program would be developed by a single developer, it was important to select a programming environment with a strong research community. Many online internet forums and source repositories were used.

Both C++ and Java satisfied the basic language requirements. These two languages were benchmarked and compared for their ability to handle singular value decompositions of huge arrays of data. C++ is a compiled language and performed quite well. Java is an interpreted language, but execution times were surprisingly quite fast. Sun Microsystems also provided Java language extensions that looked attractive: Java2D and Java3D graphics, and Java Advanced Imaging (JAI) a set of functions specific to image processing. Although these extensions were not used, there were other Java specific libraries that shortened program development cycle times. It was decided to base all programming on the Java 2 Software Development Kit (J2SDK) version 1.4. Gel, a freeware Integrated Development Environment (IDE) for Windows development of Java code was used for all programming (GExperts, 2004).

### *7.2.1 Function libraries*

ImageJ is a popular image processing program developed by Rasband (2006) at the National Institute of Health. This program first gained popularity as a Pascal program (NIH Image) but was ported to Java to embrace the needs of a wider user community. ImageJ has a comprehensive selection of 2D image processing filters, very good interactive exploratory functionality, and most important, is open source software which can be easily added to. The original plan was to develop a Multivariate Image Analysis (MIA) add-in for to ImageJ to extend its functionality. The increasing complexity of JIMIA mandated that it become a

stand-alone executable program; however the file reading class *ImagePlus*, the windowing classes *ImageWindow*, *ImageCanvas*, and *WindowMgr*, and the Region of Interest related classes *ROI, FreehandROI, PolygonROI*, and *ShapeROI* were copied from ImageJ to provide the basic mouse driven interaction.

Basic numerical processing functions for matrix manipulation were needed for JIMIA. A basic linear algebra numerical processing library (JAMA, 2005) was developed jointly by the MathWorks and NIST, based on the proven algorithms in EISPACK and LINPACK. Two classes, *SingularValueDecomposition* and *Matrix*, were used from this open source Java project.

A charting library was also needed to provide basic plotting functionality. Line charts and scatter plots of PCA loadings were provided by classes from JFreeChart (2006).

The ImagePlus file reader class from ImageJ provided reading of multi-channel (RGB) image files with BMP, JPG, PNG, or TIF file formats, and single channel GIF images. The MatrixNIR hyperspectral images acquired for research purposes in this thesis used a proprietary file format. A file reader was implemented in Java to permit reading these files directly by JIMIA.

### *7.2.2 User interface*

The basic JIMIA user interface is based on a desktop architecture. A single window (desktop) is used to contain an unlimited set of sub-windows used for display of images or score plots. A menu and toolbar provides basic navigation and feature selection. Figure 24 shows these basic design elements. The desktop window was designed for a fixed screen resolution of 1024 x 768 pixels.

The toolbar contains the following icons:

- Brushing and ROI selection including rectangular, polygon or freehand. Complex ROIs can be achieved by OR-ing different ROIs together.
- Zooming and scrolling of both images and score plots.
- Brush size selection including single pixel, 5 pt cross, and 3x3 or 5x5 pixel squares.
- Image windows containing either gray scale or false color combinations of wavelength channels, score images, or residual images.
- Score plots of any I,J pair of score vectors
- A brushing layer manager that allows display of multiple sample selection layers in a choice of colors.
- Listing of PCA model eigenvalues
- Line or scatter plotting of loading vectors
- Creation of sample subset selection masks for generation of local PCA models
- Printing and save to file functions were not implemented.

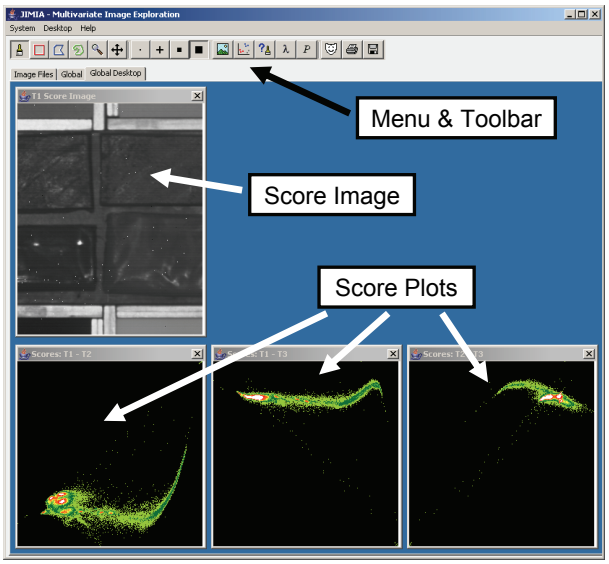


Figure 24 The basic design elements of the JIMIA desktop.

## 7.3 Application

Description of the full functionality of JIMIA is beyond the scope of this thesis; however the following set of figures demonstrates the basic functionality, based on a hyperspectral image of four cheeses. Figure 24 shows the default set of display windows: the $\mathbf{t}_1$ score image, and $\mathbf{t}_1$-$\mathbf{t}_2$, $\mathbf{t}_1$-$\mathbf{t}_3$, and $\mathbf{t}_2$-$\mathbf{t}_3$ score plots. As explained in section 3.2.2, the global PCA model is strongly distorted by the spectra corresponding to the internal standards and frame hardware. Restricting the analysis to a smaller rectangular ROI is demonstrated by the local PCA model represented in Figure 25. Unfortunately JIMIA does not detect and remove the few spectral outliers that distort the axes of the autoscaled score plots.

Figure 25 also demonstrates the first application of brushing: a small set of pixels has been selected in the score image. The locations of these sample points are highlighted in the corresponding score plots. This helps to locate the score plot regions that indicate this kind of spectral content. The inverse brushing process is demonstrated in Figure 26. Here sample points are selected in the $\mathbf{t}_1$-$\mathbf{t}_2$ score plot, and subsequently highlighted in the other score plots and score image. In this example case background pixels surrounding the cheeses have been selected based on their spectral scores.

This sample selection process can be determined by resizable and movable ROIs (rectangular, polygon, freehand, or any combination of these). For an immediate interactive mode the brush cursor allows immediate selection of only the sample points within a predefined area adjacent to the cursor. These sample points are

then highlighted immediately in all windows, providing a real-time sense of simultaneous spatial and spectral analysis.

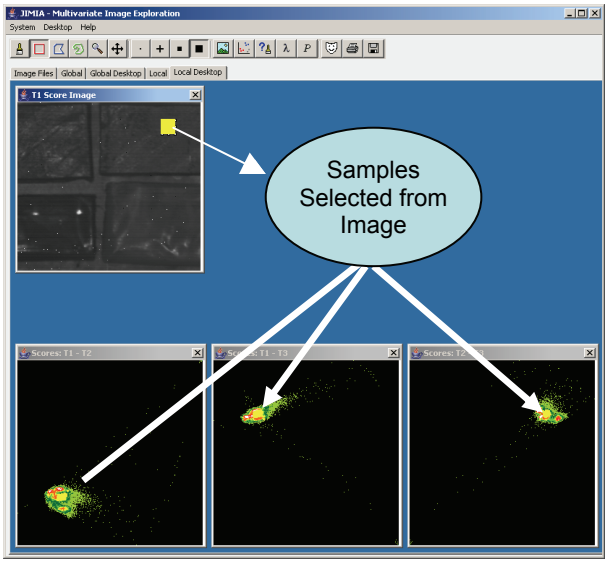


Figure 25 Image brushing. Sample pixels selected in the score image are mapped and highlighted in the score plots.

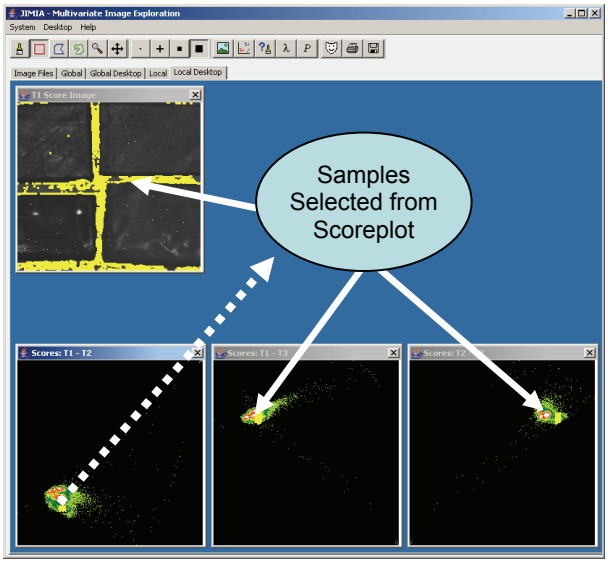


Figure 26 A set of samples is selected in the $\mathbf{t}_1$-$\mathbf{t}_2$ score plot. These samples are in turn highlighted in the other score plots and score image.

Multiple layers of highlighting can be created using different colors, allowing different classifications of spectra to be expressed. This is indicated in Figure 27. Here three different sample classes have been selected from the $\mathbf{t}_1$-$\mathbf{t}_2$ score plot. These classes map to the different cheese types, as indicated in the score image.

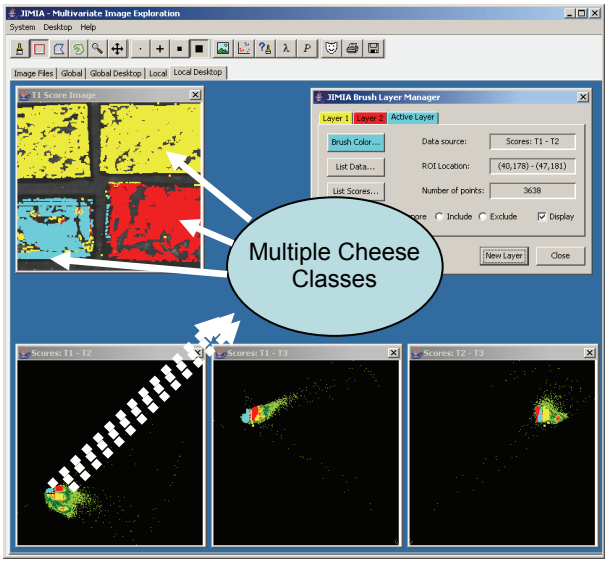


Figure 27 Multiple classes of sample points can be selected and expressed as different colors using the brush layer manager.

Sets of points selected by different brushing layers can be added or removed, to create new pixel masks for use in computing new local PCA models. The mask created by adding the three cheese areas has been indicated in Figure 28. The current implementation of JIMIA allows this mask to be exported for use in other software programs.

The JIMIA program demonstrates the utility of interactive exploration of hyperspectral image data. Additional enhancements for spatial filtering as well as inclusion of regression functionality has yet to be implemented.

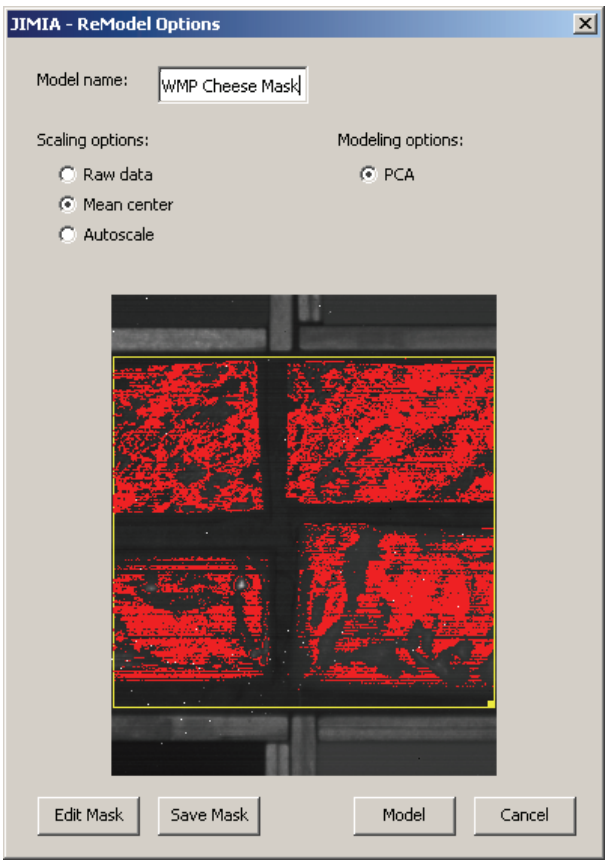


Figure 28 A new mask can be created from adding or subtracting different brushing layers. This mask can then be used to select a new pixel subset for computing new local PCA models.

# 8.0 Applications

Section 4 of this thesis described the use of external and internal standard reference materials for the transformation and standardization of hyperspectral NIR image reflectance values. Detection and removal of outliers was explained in section 5. The basis for further transformations and corrections to spectral data was expressed in section 6. This section of the thesis details the results of combining all these processes on the multivariate analysis of hyperspectral images. General conclusions and additional details relating to the four research publications are first presented, followed by several additional applications.

## 8.1 Concentration predictions

Paper I provided the theoretical foundation for the transformation of instrument signal values into percent reflectance. Papers II and III describe the results of the application of these reflectance transformations combined with spectral preprocessing treatments on the PLS regression model predictions of constituent concentrations. Both the gravimetric Tri-Mix mixture and cheese sample images were examined. Calibration models and predictions were obtained from single mean spectra from each image, effectively treating the hyperspectral imaging system as an integrating instrument. The resulting calibration models were subsequently also applied to the individual spectra contained within spatial Region of Interest (ROI) image areas. These results were summarized as histograms indicating the range and distribution of concentration predictions as well as concentration image maps showing the spatial distribution of prediction values. Both approaches demonstrate the enhanced information content available from hyperspectral imaging.

Figure 29 (Paper III) indicates the histogram and spatial mapping of the predicted concentrations of citric acid in one of the sample Tri-Mix mixtures. In this example four component PLS calibration models produced concentration prediction values with mean values very near the expected concentration of 33.4%. The width of the histogram distributions revealed a considerable variation in concentration predictions at individual pixel locations. The application of a first derivative pretreatment significantly decreased the variance in predicted values as indicated by the narrowing of the bottom histogram. A change in granularity or texture in the concentration map due to the application of the spectral pretreatment is also evident.

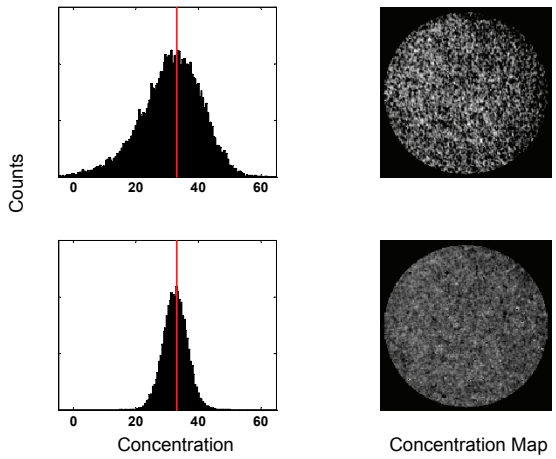


Figure 29 Mixture predictions for citric acid with an expected concentration of 33.4%. Histograms and concentration maps are for 20,000 spectra selected with a circular ROI mask with an 80 pixel radius. Prediction results are for absorbance (top) and first derivative spectra (bottom).

Unfortunately in most cases, signal variances due to problems with sample preparation and presentation to the imaging system overshadowed the variance reduction available from the application of other spectral pretreatments. Valid PLS models were obtained in all cases with similar prediction errors. This prevented detection of any clear trend in improvement of results based on pretreatment type.

## 8.2 Particle size scattering effects

The analysis of the particle size sample set is detailed in Paper IV. The objective of preprocessing was to find a transformation which collapsed the measured spectra towards a common spectrum, i.e. removing the effect of particle size scattering effects. The effects of spectral pretreatments on the mean spectra of hyperspectral images from each of the nine sugar size fractions are indicated in Figure 30. The mean spectra from the set of each of these transformed spectra were used as target spectra for the plots in Figure 31, where the individual fraction spectra are plotted vs. the target spectra at each wavelength.

The particle size dependent scatter effect was clearly evident in the reflectance spectra, as well as the absorbance, Kubelka-Munk, and 1-Norm transformed spectra. Trends from small to larger particle sizes are indicated by the arrows in Figure 30 and Figure 31. The first and second derivative transforms indicate a significant reduction in size dependency; however wavelength bands remain that show strong correlations with particle size. The MSC and especially PMSC transforms appear to remove most of the particle size dependencies.

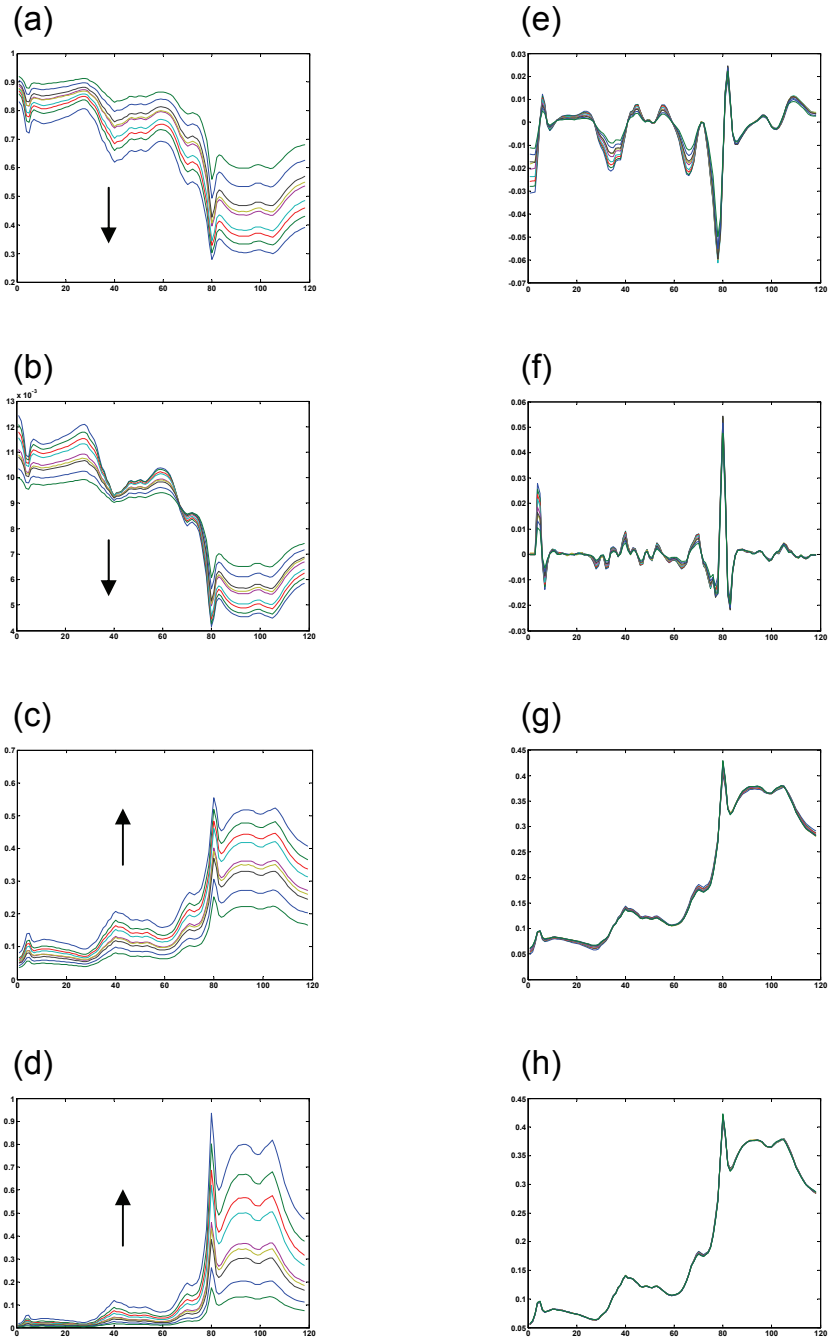


Figure 30 The effect of spectral pretreatments on the mean spectra of the nine size fractions of sugar: (a) reflectance, (b) 1-Norm, (c) absorbance, (d) Kubelka-Munk, (e) first derivative, (f) second derivative, (g) MSC of absorbance, and (h) PMSC of absorbance. The arrow indicates increasing particle size.

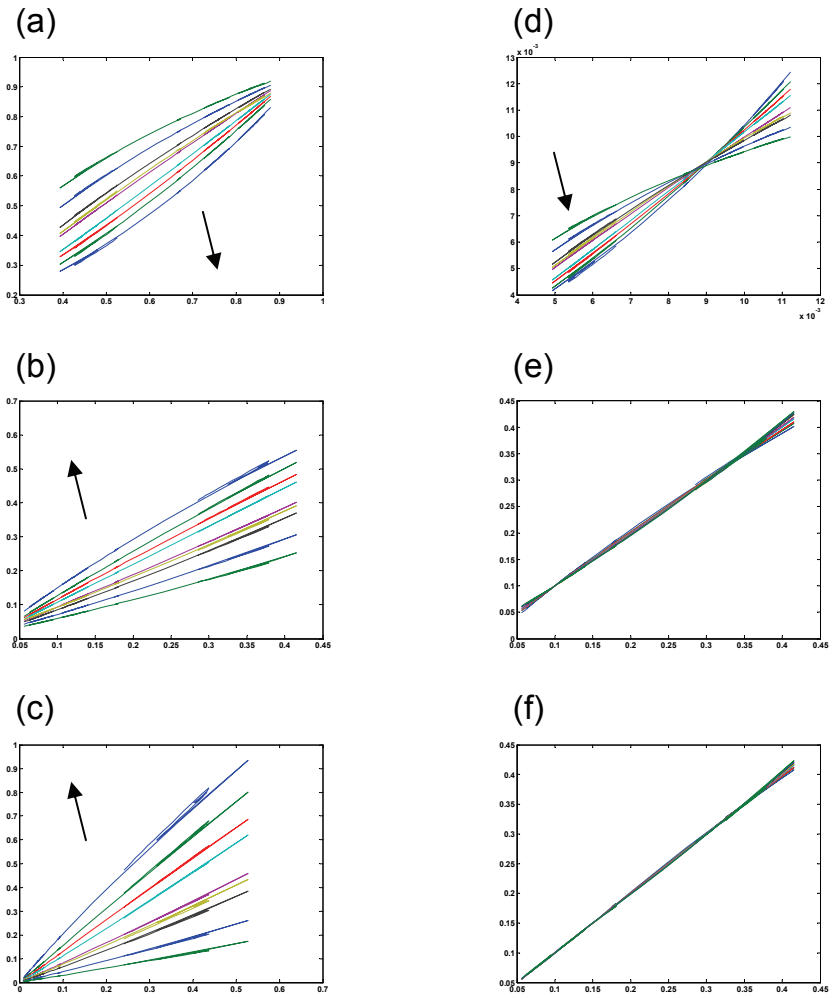


Figure 31 Plots of mean spectra for sugar size fractions vs. mean target spectrum, for various spectral pretreatments. (a) reflectance, (b) absorbance, (c) Kubelka-Munk, (d) 1-Norm, (e) MSC of absorbance, and (f) PMSC of absorbance. The arrow indicates increasing particle size.

Figure 32 presents score plots of 250 randomly selected spectra from the hyperspectral image of each of the first eight size fractions for both salt and sugar. PCA was performed after mean centering the entire combined data set. The $\mathbf{t}_1$-$\mathbf{t}_2$ score plots have been autoscaled to fit the range in computed score values. The variance of spectra from the smallest size particle fractions severely distorted the axis scaling and these fractions were therefore excluded from the plots. The trends in size dependencies are evident in all plots except the MSC and PMSC corrected data. Note that in these last two plots, the clusters with greatest variance are now salt spectra. The higher relative noise levels of the low absorbing salt spectra now have become dominant.

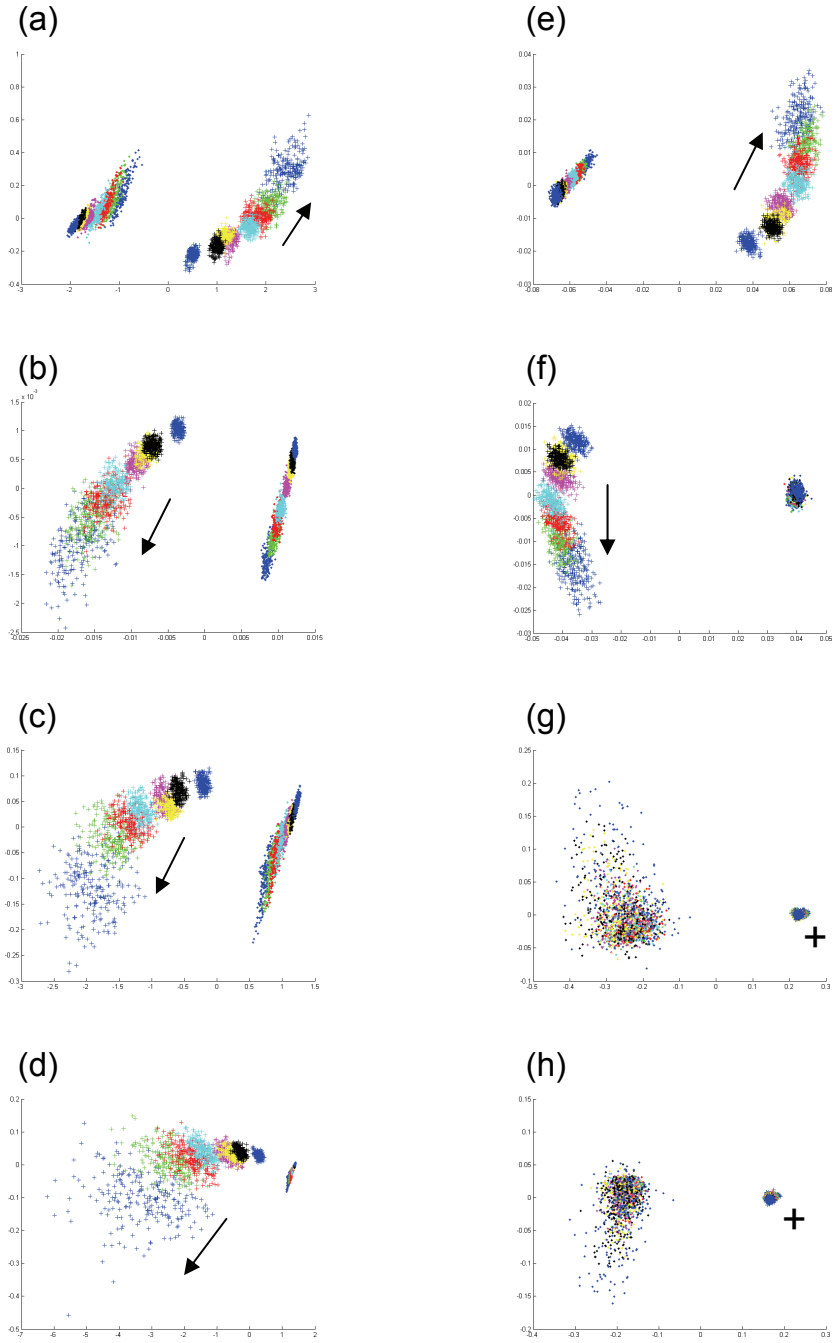


Figure 32 The effect of spectral pretreatments on all spectra of the nine size fractions of both sugar and salt: (a) reflectance, (b) 1-Norm, (c) absorbance, (d) Kubelka-Munk, (e) first derivative, (f) second derivative, (g) MSC of absorbance, and (h) PMSC of absorbance. The arrow indicates increasing sugar particle size. The plus symbol indicates the sugar cluster when size variance is not evident.

Similar score plots for second derivative spectra and SNV, MSC, or PMSC corrected second derivative spectra are presented in Figure 33. The size dependency is only evident in the original second derivative sugar spectra. Relative signal noise is clearly a dominating factor here as indicated in the larger clusters in the corrected salt spectra. This is especially evident in the PMSC corrected spectra, where the score axes have become greatly distorted by outliers.

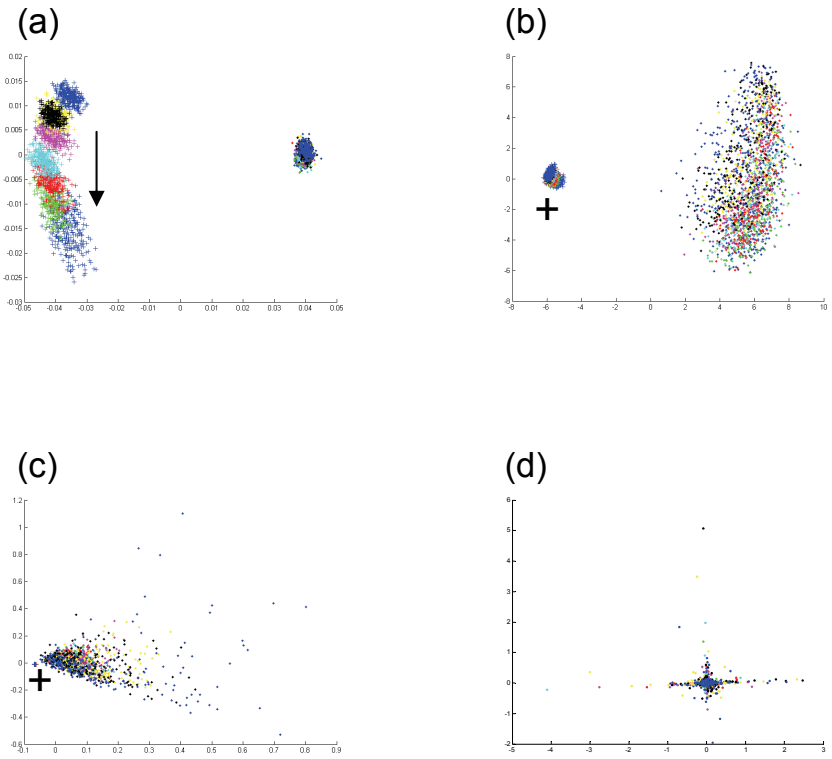


Figure 33 Score plots of all second derivative spectra of the nine size fractions of both sugar and salt (a) reflectance, (b) SNV, (c) MSC, (d) PMSC. The arrow indicates increasing sugar particle size. The plus symbol indicates the sugar cluster when size variance is not evident.

The combination of multivariate analysis tools with hyperspectral images provides a unique opportunity to explore two and three dimensional scattering effects and their relationship to particle size and shape. A more exhaustive analysis of monodisperse particles is needed to elucidate differences in chemical and physical effects of light scattering.

## 8.3 Fuel pellets

The work described in previous sections provided a solid theoretical and practical foundation for the proper acquisition and analysis of hyperspectral NIR images. The knowledge gained was applied to a more difficult application, the examination of wood fuel pellets. This alternative energy fuel is a major focus of research at SLU Unit of Biomass Technology and Chemistry, Röbäksdalen. Variations in source materials and production parameters contribute to significant differences in pellet product quality. Physical properties of pellets such as bulk density, compression strength, and moisture uptake are influenced by raw material type, moisture content, and particle size.

Hyperspectral NIR imaging was employed to investigate these differences. Raw materials, production pellets, and individually produced laboratory pellets were imaged with the hope of providing a better understanding of the production process. The spatial resolution of the imaging system and in particular the very limited depth of field available with the higher magnification optics limited the success of imaging individual pellets. Pellets are cylindrical in shape with very coarse or broken ends. Attempts were made to cut or sand pellet ends, but this was a destructive process which either fractured the pellet further or contaminated the surface with unwanted fines. A few laboratory pellets with flat ends were produced under varying conditions, but no differences were visible in the hyperspectral images acquired of these pellets.

Lower magnification optics that increased the size of the field of view (50 x 62 mm) permitted imaging of collections of pellets and was more successful. Raw materials and pellets from a series of pellet productions made at Skellefteå Kraft under varying conditions were obtained. A designed experiment of 12 different process conditions provided a variation in source material (fresh and aged pine, and spruce) as well as moisture content (8 to 12 percent moisture). The samples from the designed experiment were analyzed, both in powder form (milled and partially dried raw material) and as pellets.

Figure 34 shows an example image of pellets, including the series of internal standards used for image standardization. This image is the first principal component score image obtained after applying PCA to the complete set of pellet spectra, cleaned of significant outliers. This is a very practical application of multivariate analysis to hyperspectral imaging. The variance of all spectral channels has been combined or compressed to produce a single gray scale image with maximum signal contrast. This data compression is useful for providing summary gray scale images for use with other more traditional image processing techniques.

Figure 35 extends this exploratory or data compression technique further. Here PCA has been applied after first combining sample regions from two different hyperspectral images (fresh and aged pine) to produce an image mosaic. The image displayed is the third principal component score image. While specific chemical or

physical features have not been identified, visual inspection of this score image suggests a gross difference between pellets made from aged and fresh material.

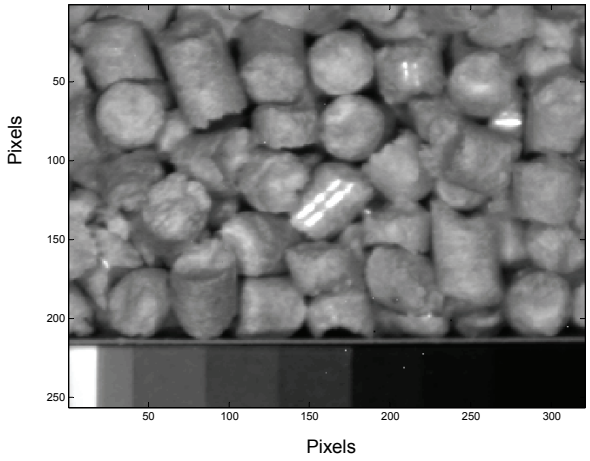


Figure 34 Principal Component Analysis (PCA) score image of pellets including internal standard reference materials for image standardization

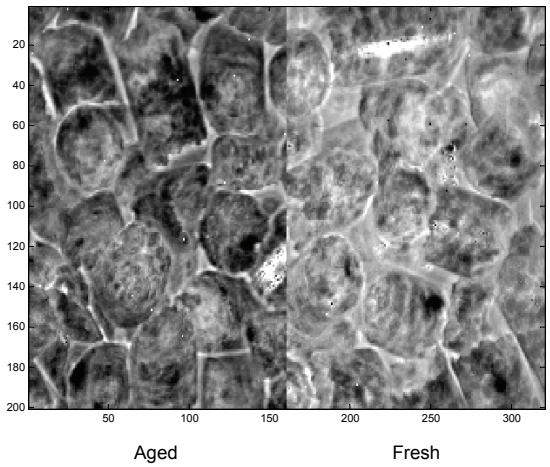


Figure 35 Third principal component score image of a mosaic image of pellets made from aged and fresh pine material.

To explore this further, PCA was applied to a set of spectra obtained from triplicate hyperspectral images of samples from all 12 pellet productions. After detecting and removing class outliers, a single average spectrum was obtained from each hyperspectral image. The 36 resulting spectra were preprocessed with a first derivative transform, followed by PCA.

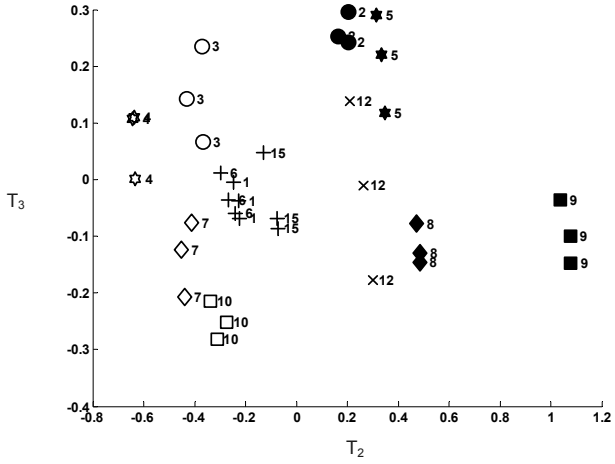


Figure 36 PCA plot of triplicate measurements of 12 pellet productions indicating separation of high and low moisture content (open and solid symbols) as well as raw material compositions (fresh pine – circle, aged pine – square, and intermediate blends.)

The score plot in Figure 36 represents the variation of all 36 spectra based on their second ($t_2$) and third ($t_3$) PCA scores. The solid symbols indicate low moisture production runs (8 %) and the open symbols high moisture (12%). These two classes are easily distinguished, primarily differentiated by their $t_2$ scores. NIR is highly sensitive to water bands so it might not seem unusual to see this discrimination of moisture content; however, the production moisture levels described here reflect the content in the dry material before pellet production. Pellet formation occurs under extreme heat and pressure, and the hyperspectral imaging was performed six months after actual pellet production. It is not known whether the observed scoreplot class separation was due to differences in water spectra, or some other byproduct of the production process.

The symbol shapes represent the composition of source material: circle is 100 percent fresh pine, square is 100 percent aged pine. The star, cross, and diamond symbols represent blends with 10 to 20 percent spruce. The plus symbols indicate a center point mixture that was repeated in three of the 12 pellet productions. The trend of blends from pure fresh to pure aged pine is also seen in Figure 2. Similar results were obtained in the hyperspectral image analysis of the dried raw source materials as well.

This analysis demonstrates the utility of hyperspectral NIR image analysis coupled with multivariate exploratory tools. Without understanding the physical intricacies of the actual processing of the pellets, basic properties could still be distinguished. To identify differences between fresh and aged starting material in such a complex matrix is quite remarkable. Further research is needed to exploit these differences.

## 8.4 Other considerations

One of the fastest and more powerful ways of extracting information from numerical data is to transform it into graphical objects. Line plots, bar charts, histograms, or scatter plots are just some of the ways of representing data that enable immediate visual interpretation. But what if the data is visual to begin with? What additional transformations can be performed to enhance or extend the normal visualization processes? The following are a few examples of simple tricks or data manipulations that assist the interpretation of hyperspectral images to better understand samples or the imaging system.

### *8.4.1 Enhancements with color*

Color itself can be used to convey additional information. This was demonstrated in the colorization of the score plots in Figure 8 and Figure 23. In this case, color indicates the relative abundance of data with identical pairs of score values; cool colors indicate relatively few counts while hot colors indicate high count levels.

Color can also be used to create false color images; Three gray scale images can be mapped to levels of red, green, and blue respectively and overlaid to form a composite color image. Different relative data values within the red, green, and blue channels will combine to yield different false colors. This technique can be applied to sets of individual wavelength channels, score images, residual images, or scaled prediction levels of multiple constituents. Figure 7 in Paper II demonstrates the benefit of combining prediction images of the three ingredients, sugar, citric acid, and salicylic acid, in one of the TriMix sample images. Figure 9 in Paper II and Figure X below demonstrate the use of color in representing the fat, protein, and carbohydrate content of cheese.

### *8.4.2 Incorporating additional latent variables*

Normally the number of latent variables to include in a PLS calibration model is selected to minimize measured residual errors determined from a test set of data. The variance contained in additional components may not be correlated to the dependent Y-block variables, however it may contain useful information. Figure 37 shows the color mapped predictions for fat, protein, and carbohydrate (red, green, and blue channels) for four cheeses. The top row of images represent the predictions obtained from a four component PLS model based on absorbance spectra. The middle and bottom rows represent the results from PLS models utilizing first derivative spectra for two and nine components respectively.

The colored rectangles below each image indicate the ‘target color’ expected for each cheese. In all cases the general prediction levels match quite well for this range of four cheeses; however, the detailed features (or lack of) in parallel images reveals additional information. Physical factors such as the cutting marks in cheese b are removed by the first derivative transform. Texture effects in the other cheeses, possibly due to specular reflection, are also removed by the derivative transform. But the nine component prediction images in the bottom row indicate a new phenomena: coinciding circular ring like features in all

images. This is an artifact created by spinning the SRM reflectance tiles used for the reflection transform process. In this case diagnostic information regarding the imaging system has been obtained.

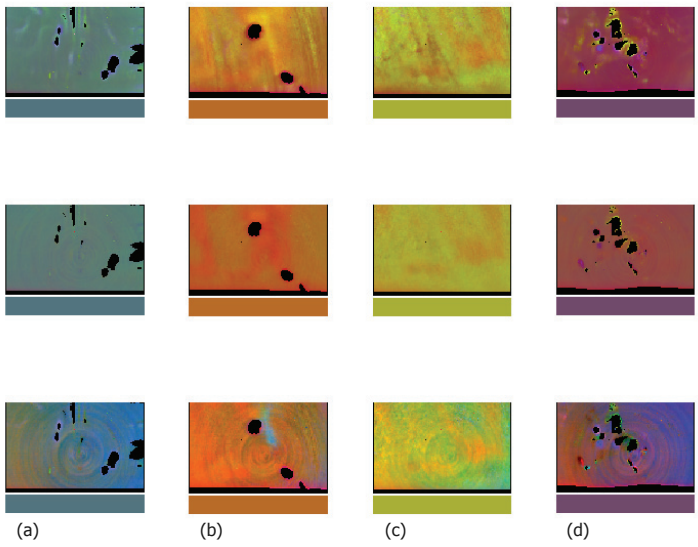


Figure 37 False color images indicate the fat (red), protein (green), and carbohydrate (blue) content of four cheeses (a – d). Predictions are based on a four component PLS model of absorbance spectra (top), or PLS models of first derivative spectra with two components (middle) or nine components (bottom). The circular ring feature in the bottom row is an artifact from the spinning tiles used for the instrument signal to reflectance transformation.

### *8.4.3 Sample masks*

Additional system information was determined by purposely modifying image acquisition conditions. A simple visual mask was created to help assess the depth of penetration of scattered light in the imaging system. An imaging background was created by partially covering the 99 percent reflectance SRM with a piece of silicon carbide sandpaper and cover glass. This resulted in a field of view consisting of two large rectangular areas with measured average reflectance values of 7.8 and 98.3 percent. Glass microscope slides were added and positioned to create a sample frame 1 mm thick. This area was filled with a uniform layer of a mixture of starch, cellulose, acetylsalicylic acid, ascorbic acid, and caffeine and imaged.

Spectra from two rectangular 100 x 200 pixel ROIs in front of the *white* and *black* reflectance backgrounds were analyzed. Median spectra were computed from each set of 20,000 spectra to minimize effects of outliers and are displayed in Figure 38. Two wavelength channels were selected for further analysis based on

these spectra: the 1086 nm channel with relatively high defuse reflectance values, and the absorbance maximum at 1452 nm. Histograms of the 20,000 values for each ROI are displayed in Figure 39.

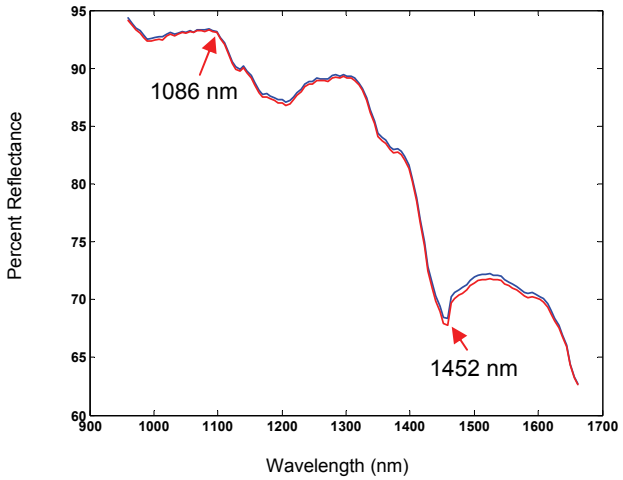


Figure 38 Median spectra from 20,000 spectra measured with a white or black reflectance background.

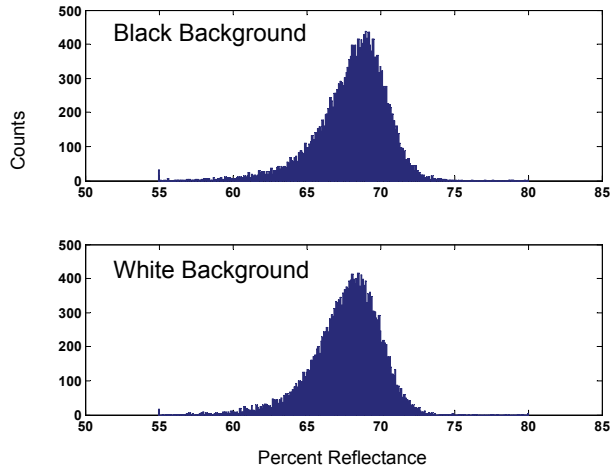


Figure 39 Histograms of 20,000 spectra measured with a white or black reflectance background.

Score plots computed from a global PCA model of the combined set of 40,000 spectra were also made. Based on the score plots, histograms and median spectra plots, no discrimination was possible between the two classes of spectra with the white or black background. This indicated that the 1 mm sample thickness was effectively blocking the extreme differences in background created by the simple visual mask. This approach could be extended with other sample thicknesses or sample materials to characterize the actual light penetration levels.

# 9.0 Conclusions

Hyperspectral NIR images contain an enormous amount of data, but to effectively extract meaningful information in an efficient way, a combination of chemometric and exploratory visualization tools are needed. While the sheer volume of data may cause some computational burdens, it provides unique advantages as well. First, because there is so much data, it may be easily justifiable to set aside suspicious data for further examination, or throw it out completely. Second, sample data space can be specifically allocated for inclusion of internal standard or Standard Reference Materials (SRMs). Both of these approaches lead to improving the consistency or quality of the remaining data. Because there is no spatial averaging effect as with integrating spectrometers, it is important to minimize any individual spectral disturbances. It is especially important to ensure that the data is conditioned properly to avoid misuse and erroneous conclusions reached by automated processing.

Several general conclusions can be made about this data cleansing:

- Image data should be acquired and saved as raw instrument counts. The transformation of this data into percent reflectance units should be based on transform functions computed from the second order regression models of a series of images of SRMs with known reflectance values. Retaining the original image data in raw instrument count format allows for optimal examination in case of instrument malfunction.

- Internal standards should be incorporated into the image field of view to allow instrument standardization. As with the raw data format, the internal standards are invaluable in case of instrument malfunction.

- Outliers should be detected and removed. A combination of detection techniques is needed, that should include excessive signal or threshold detection, as well as tests based on the Euclidean distances and angle cosines between each spectrum and a single mean target spectrum.

- Depending on instrument or sample instabilities, spectral pretreatments may be applied to reduce or eliminate instable signal contributions.

Three general types of analysis tools should be employed with the cleansed or conditioned hyperspectral data:

- Univariate statistical techniques are still important for the analysis of the large sets of spectra and prediction data. Accurate population statistics estimates such as mean and standard deviation can be computed.

- Multivariate statistical techniques are essential to take advantage of correlations between spectra and variables. But because of the sensitivity

to noise, regression based calibration models such as PLS should be based on mean spectra from ROIs representing classes of spectra with known response variables.

- Visualization tools must be used to facilitate exploration of the data. These tools include histograms, colorized score plots, and both gray scale and false color images representing score values, prediction values, and residuals.

Properly cleansed and pretreated data from hyperspectral images can yield robust calibration models that in turn can be used to provide accurate sample information. Bulk concentration values can be determined as with integrating spectrometer systems. But more importantly, concentrations can be combined with spatial information to produce concentration maps. Concentrations can also be displayed as histograms or concentration profiles. Both techniques reveal uniformity or homogeneity information that would be very time consuming and costly to acquire with an integrating spectrometer.

# 10.0 Future research

The combination of spectral and spatial information afforded in hyperspectral imaging systems provides a powerful research tool. The acquisition of hyperspectral images provides extensive data sets for testing of new algorithms and studying of physical / chemical phenomena.

But there are difficulties with instrumentation and sample presentation that need to be addressed. Faster instrumentation is needed to image biological materials or other samples which are heat sensitive. Better control of sample presentation inconsistencies is needed. Particle size, packing force, light penetration depth, and static charge effects, are just a few parameters that need to be examined.

Standard reference materials need to be produced which address the sample homogeneity issues at higher and higher magnifications. SRMs are needed for both intensity and wavelength calibration. Because multivariate statistics tools are employed, SRMs don't need to have a flat uniform spectrum as is encouraged with the Spectralon material. It is only necessary to have a set of standards that in total, span all expected sample reflectance values at all wavelengths; i.e. the two SRMs that provide the minimum and maximum reflectance values at one wavelength need not be the same two at any other wavelength. A golden rule of chemometrics should be applied here: "Math is cheaper than physics."

Once the instrumentation and sampling issues are addressed, carefully designed experiments should be performed to further investigate the effects of spectral pretreatments on chemical and physical interactions such as the diffuse scattering of light in powders, thin films and colloidal suspensions. Samples with precisely known particle sizes, shapes, and spectral features are needed.

Other general spectroscopic issues should be examined that were totally ignored in the current research. Variable or wavelength selection needs to be addressed. Alternative regression techniques and other spectral pretreatments should be examined as well.

Other configurations of hyperspectral imaging systems offer new opportunities as well. For example, the line scan camera with dispersing optics allows operation in a continuous mode. Additional adaptations to existing analysis and visualization tools are needed to exploit this continuous operation mode.

One aspect of hyperspectral image analysis needs extensive research: the question of "How can the correlations between data from adjacent pixels and adjacent wavelengths be fully exploited?" This issue has been slightly addressed in the remote sensing field, but mostly ignored in the chemical laboratory. For example, all spectral processing that was accomplished in the research performed for this thesis utilized an *unfolded* hyperspectral image. With the exception of pixel indices used for concentration and residual mapping images, the spatial positioning information associated with each spectrum was totally ignored. That

is to say, the spectra in the unfolded X-block matrix could be randomly re-ordered, with absolutely no effect on techniques such as PCA or PLS. A similar claim can be made for the wavelength variables: the individual channels in the original hyperspectral image could be re-ordered, and no change in PCA results would occur, other than the shuffling from the re-ordering step.

And yet there are very strong correlations between the neighbors in a hyperspectral image. How can this linkage information between adjacent pixels or adjacent wavelengths be captured to improve calibration and prediction or classification processes? Only when these linkages are completely exploited, will the rich information content of hyperspectral images become fully unlocked.

# 11.0 Acknowledgements

A PhD program that has spanned over two decades includes the support of many individuals and organizations. I am indebted to the many people that have helped guide my travels along this long and meandering path. There are of course, a few people I am especially grateful to.

In particular, this thesis is dedicated to my father: an amazing scientist, incredible teacher, and loving dad. Throughout the years (and decades) you've been there, nurturing my passion to explore the world around me, ask questions, and seek the *new*. You taught me how to seek out simple examples, to comprehend the most complicated theories. You encouraged me to take things apart; to seek out the how and why they worked and eventually, even to learn how put things back together! And as if that wasn't enough, simultaneously you fueled my passion for math. I wouldn't be where I am today without your support. Thank you.

I also want to acknowledge the continual support and love from my son, Jeff. Being both a student and father is not an easy role, but over the years your never-ending flow of questions have helped me to open my eyes a little more, reach a little further, or probe a little deeper. On a scale of 10, I would say life with you has been at least a 9.24. Thanks kiddo.

I've had a few other great teachers in my life I want to express my gratitude to:

Calvin Gentle was my high school calculus teacher. He had a rare enthusiasm for math that I have never forgotten. Probably more than anything else, he taught me the importance of not memorizing things (equations), but to learn how to derive them instead. With a solid foundation and understanding, you can build anything…

Ed Lingafelter was an old family friend who also happened to be a great p-chem teacher while I was first attending the University of Washington. He was also a great friend and my employer for several years. I spent days and weeks, keypunching boxes of data cards to be read by huge mainframe computers and scanned for patterns. Ed launched my career and interest in using computers to explore for chemical information. If only we had had some chemometrics tools. But that was just around the corner, literally…

Not more than 30 meters down the hall was a guy developing some other computer software known as *Arthur*. During my previous work period with Ed I had never met this guy, but five years later I was invited back to the UW, to begin a PhD program with Bruce Kowalski. Wow, what an impact. I bounced in and out of the PhD program for ten years. I didn't complete the degree, but thanks to Bruce, I received the most comprehensive training in chemometrics imaginable. There was great support from not only within the research group, but also a long list of names of visiting scholars (and ski bums); however the center point of this

community was Bruce. There is no doubt in my mind that Bruce deserves the title 'father of chemometrics'. Much of my inspiration comes from the group meetings, and the one-on-one research meetings with Bruce that planted so many seeds in my mind. Not completing the PhD at that time has been a huge disappointment for me, but I haven't given up the quest! This thesis is dedicated to you too, Bruce. Thank you.

Ok, fast forward 15 years past a few programming bug cycles. Kim Esbensen and Rasmus Bro came into my life when I needed them most. A four day workshop course they taught in 2001 peaked my interest in something called Multivariate Image Analysis (Kim) and thinking in more ways than… well, just more *ways* (Rasmus). I am especially grateful to Kim for launching my career back into the chemometrics world, and back on course in the PhD program. I took a huge risk moving to Europe to pursue this dream, and Kim helped smooth the way. Thank you, sir.

Throughout my Scandinavian years, I have continued to bounce around multi-way topics and hyperspectral image analysis with Rasmus. Thanks for the continued support with both the research and the PhD road. There is so much yet to be done, we'll get there yet!

And now a special man I am very, very, indebted to: Paul Geladi has been a great advisor, teacher, translator, and friend to me. It's not always easy being a stranger in a strange land. Paul has constantly been there for me, providing me with guidance and instruction when I need it and freedom and space when I don't. Thanks to Paul, I have never had to worry about the funding battles, whether for salary, instrumentation, or travel. He has been patient with my writing, and engaged me in numerous discussions ranging from NIR spectroscopy and PLS, to Swedish culture and music. He's even been my personal car attendant, clearing it of snow while I'm away! Paul, You have never faltered in your endless support of me. I could never have achieved this degree without you. Thanks.

There are many others who have helped me to make this thesis possible:

Many thanks to another good friend and technical colleague, Hans Grahn, for many enlightening MIA discussions and good times. Keep that Pontiac rollin…

Thanks also to Agnar Höskuldsson for the theoretical and practical statistics discussions and to Torbjörn Lestander and Michael Sjöström for their guidance and support with the MatrixNIR camera.

My very first days of hyperspectral imaging were with Bob Cogdill from Duquesne University, Pittsburgh, PA, USA. Thanks for the instrument standardization and practical imaging discussions here on our first visit to Umeå.

The internal standards used in my research were complimentary gifts from Jan Sjögren from TestElek Svenska AB, and Steve Choquette, National Institute of Standards and Technology. Your contributions are truly appreciated.

Additional help with samples and imaging work were provided by Olga Kolomiets, University of Essen, Germany, and Fabien Deswarte, University of York, UK. Carina Jonsson, SLU provided some last minute fat and protein analysis of cheese. Thank you.

Many instrumentation and general spectroscopy questions were answered by Bo Karlberg and Anders Nilsson. Thanks, it's been good to work with you.

I received specialized fuel pellet samples and had numerous related discussions with Rolf Gref, SLU. Thanks Rolf, for sharing your knowledge of this simple yet complex production process. And a very special thanks for the last minute review of the thesis manuscript.

There have been so many people at the Unit of Biomass Technology and Chemistry, SLU that have contributed to my survival in the far north of Sweden. You really have made my life enjoyable, even if I struggle with fika. Thanks for putting up with my even stranger American work habits.

Special thanks to Maj-Brith Ragnarsson for all your help with general administrative matters, especially making sure that I was always a legal Swedish alien. I'm also grateful to Robert Samuelsson for his assistance with all the PhD program administration issues, and providing me with more fuel pellets than I could ever dream of.

I'm very appreciative of the funding support from Rolf Olsson and the SLUP project. Financial support was also provided by NIRCE, an EU Unizon-Kwarken project. Instrumentation funding was provided by Kempestiftelserna grant SMK-2062.

And finally some very special people that have provided me with endless support through thick and thin must be acknowledged:

Glenn Gillette, thanks for sharing your need for speed, and your love of kick-ass rock-n-roll. I would never have survived the countless days and nights of programming without your never ending supply of music CDs and DVDs. Pete Townshend was right: "Long live rock, I need it every night".

Thank you Anne 'Doe' Burger Cole for helping me to find the courage to take this leap and chase after my dreams. It would have been much harder to jump by myself. You will always be my sunshine. And thanks also for the meticulous thesis proofreading, that which however, may have been incomplete.

Aelita, Thank you for encouraging me to think outside the box; in fact, thank you for showing me there is no box! It seems a hyperspectral image is nothing more than multi-dimensional light-dancing. Thanks for teaching me a few new dance steps… Let the real show begin!

And last but not least, I cannot forget my four-legged buddies, Molly, Buster, and Kody. Your constant companionship and enthusiasm for playtime has helped to smooth the ups and downs we humans seem to inflict on life. Thank you so much! Now, where's the Frisbee?

# 12.0 References


Abney, W. & Festing, E. 1881. On the influence of the atomic grouping in the molecules of organic bodies on their absorption in the infra-red region of the spectrum. *Philosophical Transactions of the Royal Society* **172**; 887-918.

Aikio, M. 2001. *Hyperspectral Prism-Grating-Prism Imaging Spectrograph*. PhD Thesis 435, VTT Publications, University of Oulu, Finland.

Antonelli, A., Cocchi, M., Fava, P., Foca, G., Franchini, C., & Manzini, D. 2004. Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet based classification algorithm. *Analytica Chimica Acta* **515**; 3-13.

Artyushkova, K. & Fulghum, J. 2001. Identification of chemical components in XPS spectra and images using multivariate statistical analysis methods. *Journal of Electron Spectroscopy and Related Phenomena* **121**; 33-55.

Artyushkova, K & Fulghum, J. 2002. Multivariate image analysis methods applied to XPS imaging data sets. *Surface and Interface Analysis* **33**; 185-195.

Attas, M., Cloutis, E., Collins, C., Goltz, D., Majzels, C., & Mansfield, J. 2003. Near-infrared spectroscopic imaging in art conservation: investigation of drawing constituents. *Journal of Cultural Heritage* **4**; 127-136.

Bacci, M., Casini, A., Cucci, C., Muzzi, A., & Porcinai, S. 2005. A study on a set of drawings by Parmigianino: integration of art-historical analysis with imaging spectroscopy. *Journal of Cultural Heritage* **6**; 329-336.

Barnes, P. & Hsia, J. 1995. 45°/0° Reflectance Factors of Pressed Polytetrafluoroethylene (PTFE) Powder. *NIST Technical Note 1413* NIST, Gaithersburg, MD.

Barnes, R., Dhanoa, M., & Lister, S. 1989. Standard normal variate transformation and detrending of near infrared diffuse reflectance. *Applied Spectroscopy* **43**; 772-777.

Baronti, S., Casini, A., Lotti, F., & Porcinai, S. 1997. Principal component analysis of visible and near-infrared multispectral images of works of art. *Chemometrics and Intelligent Laboratory Systems* **39**; 103-114.

Barton, F. 2004. Progress in near infrared spectroscopy: the people, the instrumentation, the applications. In: *Near Infrared Spectroscopy: Proceedings of the 11th International Conference*. (Eds. A.M.C. Davies & A. Garrido-Varo). NIR Publications, Chichester, UK. 13-18.

Beebe, K., Pell, R., & Seasholtz, M. 1998. *Chemometrics, A Practical Guide* John Wiley & Sons, New York, USA.

Beltrami, M. 1873. Sulle funzioni bilineari. *Giornale di Matematiche ad Uso degli Studenti Delle Universita* **11**; 98-106.

Borregaard, T., Nielsen, H., Norgaard, L., & Have, H. 2000. Crop-weed discrimination by line imaging spectroscopy. *Journal of Agricultural Engineering* **75**; 389-400.

Brown, S. 1990. Review of Chemometrics. *Analytical Chemistry* **62**; 84R-101R.

Chiang, S., Chang, C., & Ginsberg, I. 2000. Unsupervised hyperspectral image analysis using independent component analysis (ICA). In: *Proceedings of the 2000 IEEE International Geoscience and Remote Sensing Symposium Volume 7*. IEEE, New York, USA. 3136-3138.

Choquette, S. 2000. Near-infrared wavelength characterization of eight absorbance peaks of polystyrene in the 10000 $cm^{-1}$ to 4000 $cm^{-1}$ region. *Report of Analysis* NIST 839.04-00-224.

Choquette, S., Duewer, D., Hanssen, L., & Early, E. 2005. Standard reference material 2036 near-infrared reflection wavelength standard. *Applied Spectroscopy* **59**; 496-504.

Choquette, S., O'Neal, L., & Duewer, D. 2003. Rare-earth glass reference materials for near-infrared spectrometry: correcting and exploiting temperature dependencies. *Analytical Chemistry* **75**; 961-966.

Choquette, S., Travis, J., Zhu, C., & Duewer, D. 2002. Wavenumber Standards for Near Infrared Spectroscopy. In: *Handbook of Vibrational Spectroscopy Volume 1*. (Eds. J.M. Chalmers & P.R. Griffiths). John Wiley & Sons Ltd, Chichester, UK.

Coblenz, W. 1905. Investigations of Infrared Spectra, Part 1. *Publication No 35*, Carnegie Institute of Washington: Washington DC. (republished under the joint sponsorship of the Coblenz Society and the Perkin-Elmer Corporation, 1962).

Dhanoa, M., Lister, S., Sanderson, R., & Barnes, R. 1994. The link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformations of NIR spectra *Journal of Near Infrared Spectroscopy* **2**; 43-47.

Ellis, J. 1929. Molecular absorption spectra of liquids below 3 μ. *Transactions of the Faraday Society* **25**; 888-898.

Esbensen, K. & Geladi, P. 1989. Strategy of multivariate image-analysis (MIA). *Chemometrics and Intelligent Laboratory Systems* **7**; 67-86.

Esbensen, K., Geladi, P., & Grahn, H. 1992. Strategies for multivariate image regression. *Chemometrics and Intelligent Laboratory Systems* **14**; 357-374.

Esbensen, K., Wold, S., & Geladi, P. 1988. Relationships between higher-order data array configurations and problem formulations in multivariate data analysis. *Journal of Chemometrics* **3**; 33-48.

Frank, I. 1987. Intermediate least squares regression method. *Chemometrics and Intelligent Laboratory Systems* **1**; 233-242.

Fuxi, G. 1992. *Optical and Spectroscopic Properties of Glass*. Springer-Verlag, New York, USA.

Geladi, P. 1995. Sampling and local models for multivariate image-analysis. *Mikrochimica Acta* **120**; 211-230.

Geladi, P. 2003. Chemometrics in spectroscopy. Part 1. Classical Chemometrics. *Spectrochimica Acta B* **58**; 767-782.

Geladi, P. & Esbensen, K. 1990. The Start and Early History of Chemometrics: Selected Interviews, Part 1. *Journal of Chemometrics* **4**; 337-354.

Geladi, P. & Esbensen, K. 1991. Regression on multivariate images: principal component regression for modelling, prediction and visual diagnostic tools. *Journal of Chemometrics* **5**; 97-111.

Geladi, P., Isaksson, H., Lindqvist, L., & Wold, S. 1988. Principal component analysis on multivariate images. *Chemometrics and Intelligent Laboratory Systems* **5**; 209-220.

Geladi, P., McDougall, D. & Martens, H. 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy* **39**; 491-500.

Geladi, P., Swerts, J., & Lindgren, F. 1994. Multiwavelength microscopic image-analysis of a piece of painted chinaware-classification and regression. *Chemometrics and Intelligent Laboratory Systems* **24**; 145-167.

Geladi, P., Wold, S., & Esbensen, K. 1986. Image analysis and chemical information in images. *Analytica Chimica Acta* **191**; 473-480.

GExperts. 2004. Home of the Java IDE Gel. http://www.gexperts.com/ (Accessed August 6, 2006).

Giesbrecht, F., McClure, W., & Hamid, A. 1981. The use of trigonometric polynomials to approximate visible and near infrared spectra of agricultural products *Applied Spectroscopy* **35**; 210-214.

Golub, G. & Reinsch, C. 1970. Singular value decomposition and least squares solution. *Numerical Mathematics* **14**; 403-420.

Grahn, H. & Saaf, J. 1992. Multivariate image regression and analysis-useful techniques for the evaluation of clinical magnetic-resonance images. *Chemometrics and Intelligent Laboratory Systems* **14**; 391-396.

Grahn, H., Szeverenyi, N., Roggenbuck, M., & Geladi, P. 1989a. Tissue discrimination in magnetic resonance imaging: a predictive multivariate approach. *Chemometrics and Intelligent Laboratory Systems* **7**; 87-93.

Grahn, H., Szeverenyi, N., Roggenbuck, M., DeLaglio, F., & Geladi, P. 1989b. Data analysis of multivariate magnetic resonance images I. A principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **5**; 311-322.

Hardy, A., MacLaurin, P., Haswell, S., de Jong, S., & Vandeginste, B. 1996. Double-case diagnostic for outliers identification. *Chemometrics and Intelligent Laboratory Systems* **34**; 117-129.

Herschel, W. 1800. Investigation of the power of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously with telescopes of large apertures and high magnifying powers. *Philosophical Transactions of the Royal Society of London* **90**; 255-283.

Hopkins, D. 2001. Derivatives in spectroscopy. *Near Infrared Analysis* **2**; 1-13.

Höskuldsson, A. 1988. PLS regression methods. *Journal of Chemometrics* **2**; 211-228.

Isaksson, T. & Kowalski, B. 1993. Piece-wise multiplicative scatter correction applied to near-infrared diffuse transmittance data from meat products. *Applied Spectroscopy* **47**; 702-709.

Isaksson, T., Yang, H., Kemeny, G., Jackson, R., Wang, Q., Alam, M., Griffiths, P. 2003.Accurate wavelength measurements of a putative standard for near-infrared diffuse reflection spectrometry. *Applied Spectroscopy* **57**; 176-185.

Jackson, J. 1991. *A User's Guide to Principal Components* Wiley, New York, USA.

JAMA. 2005. JAMA – a basic linear algebra package for Java. http://math.nist.gov/javanumerics/jama/ (Accessed August 6, 2006).

JFreeChart. 2006. JFreeChart is a free 100% Java chart library for developers. http://www.jfree.org/jfreechart/ (Accessed August 6, 2006).

Jordan, C. 1874. Mémoire sur les formes bilinéaires *Journal de Mathématiques Pures et Appliquées*, *Deuxième Série* **19**; 35-54.

Katsumoto, Y., Jiang, J., Berry, J., & Ozaki, Y. 2001. Modern pretreatment methods in NIR spectroscopy. *Near Infrared Analysis* **2**; 29-36.

Kaye, W. 1954. Near-infrared spectroscopy; A review. I. Spectral identification and analytical applications. *Spectrochimica Acta* **6**; 257-287.

Kortum, G. 1969. *Reflectance Spectroscopy*. Springer-Verlag, New York, USA.

Kubelka, P. & Munk, F. 1931. Ein Beitrag zur Optik der Far-banstriche. *Zeitschrift für Technische Physik* **12**; 593-604.

Liang, Y. & Kvalheim, O. 1996. Robust methods for multivariate analysis - A tutorial review. *Chemometrics and Intelligent Laboratory Systems* **32**; 1-10.

Lied, T. & Esbensen, K. 2001. Principles of MIR, multivariate image regression I: Regression typology and representative application studies. *Chemometrics and Intelligent Laboratory Systems* **58**; 213-226.

Lied, T., Geladi, P., & Esbensen, K. 2000. Multivariate image regression (MIR): implementation of image PLSR-first forays. *Journal of Chemometrics* **14**; 585-598.

Madden, H. 1978. Comments on smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **50**; 1383-1386.

Malinowski, E. 2002. *Factor Analysis in Chemistry, 3rd ed*. Wiley, New York, USA.

Marengo, E., Aceto, M., Robotti, E., Liparota, M., Bobbaa, M., & Panto, G. 2005. Archaeometric characterization of ancient pottery belonging to the archaeological site of Novalesa Abbey (Piedmont, Italy) by ICP-MS and spectroscopic techniques coupled to multivariate statistical tools. *Analytica Chimica Acta* **537**; 359-375.

Martens, H., Jensen, S., & Geladi, P. 1983. Multivariate linearity transformations for near-infrared spectrometry. In: *Proceedings of the Nordic Symposium on Applied Statistics.* (Ed. O. Christie). Stokkand Forlag, Stavanger. 205-233.

Martens, H., Nielsen, J., & Engelsen, S. 2003. Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry* **75**; 394-404.

Martens, H., & Næs, T. 1989. *Multivariate Calibration*. John Wiley & Son, Chichester; UK.

Martens, H. & Stark, E. 1991. Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **9**; 625-635.

Massart, D., Vandeginste, B., Deming, S., Michotte, Y., & Kaufman, L. 1988. *Chemometrics: A Textbook* Elsevier Science Publishers, Amsterdam.

Miseo, E. & Wright, N. 2003. Developing a Chemical-Imaging Camera. *the Industrial Physicist* **9**; 29-32.

Møller, S., von Frese, J. & Bro, R. 2005. Robust methods for multivariate analysis. *Journal of Chemometrics* **19**; 549-563.

Morrey, J. 1968. On determining spectral peak positions from composite spectra with a digital computer. *Analytical Chemistry* **40**; 905-914.

Murray, I. 2004. Scattered information: philosophy and practice of near infrared spectroscopy. In: *Near Infrared Spectroscopy: Proceedings of the 11th International Conference*. (Eds. A.M.C. Davies & A. Garrido-Varo). NIR Publications, Chichester, UK. 1-12.

Næs, T., Isaksson, T., Fearn, T., & Davies, T. 2002. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK.

NASA. 2006. National Aeronautics and Space Administration Remote Sensing Tutorial home page http://rst.gsfc.nasa.gov/Homepage/Homepage.html (Accessed July 15, 2006).

Norris, K. 1962. Instrumentation of infrared radiation. *Transactions of the American Society of Agricultural Engineers* **5**; 17-20.

Norris, K. & Butler, W. 1961. Techniques for obtaining absorption spectra on intact biological samples. *IRE Transactions on Biomedical Electronics* **8**; 153-157.

Norris, K. & Williams, P. 1984. 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 Chemistry* **61**; 158-165.

Osborne, B., Fearn, T., & Hindle, P. 1993. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis* 2nd Edition. Longman Scientific & Technical. Essex, England.

Park, B., Lawrence, K., Windham, W., & Smith, D. 2004. Multispectral Imaging system for fecal and ingesta detection on poultry carcasses. *Journal of Food Process Engineering* **27**; 311-327.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**; 559-572.

Pierna, J., Baeten, V., Renier, A., Cogdill, R., & Dardenne, P. 2004. Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics* **18**; 341-349.

Rasband, W. 2006. ImageJ – Image processing and analysis in Java. http://rsb.info.nih.gov/ij/ (Accessed August 6, 2006).

Sahlin, J. & Peppas, N. 1997. Near-field FTIR Imaging: A technique for Enhancing Spatial Resolution in FTIR Microscopy. *Journal of Applied Polymer Science* **63**; 103-110.

Sanchez, E. & Kowalski, B. 1988a. Tensorial Calibration: I. First-order Calibration. *Journal of Chemometrics* **2**; 247-264.

Sanchez, E. & Kowalski, B. 1988b. Tensorial Calibration: II. Second-order Calibration. *Journal of Chemometrics* **2**; 265-280.

Savitzky, A. & Golay, M. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **36**; 1627-1639.

Shenk, J. & Westerhaus, M. 1991. New standardization and calibration procedures for NIRS analytical systems. *Crop Science* **31**; 1694-1696.

Shotton, D. ed., 1993. *Electronic Light Microscopy: Techniques in Modern Biomedical Microscopy.* Wiley-Liss, New York.

Sjöström, M., Wold, S., Lindberg, W., Persson, J., & Martens, H. 1983 A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables. *Analytica Chimica Acta* **150**; 61-70.

Smilde, A., Bro, R., & Geladi, P. 2004. *Multi-way Analysis*. John Wiley & Sons, Chichester, England.

SpecIm. 2006. Spectral Imaging. http://www.specim.fi (Accessed July 15, 2006).

Stark, E. & Luchter, K. 2004. Diversity in NIR instrumentation. In: *Near Infrared Spectroscopy: Proceedings of the 11th International Conference*. (Eds. A.M.C. Davies & A. Garrido-Varo). NIR Publications, Chichester, UK. 55-66.

Steiner, J., Termonia, Y., & Deltour, J. 1972. Comments on smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* **44**; 1906-1909.

Švedas, V. 2004. Spectral pretreatment for diffuse transmittance linearity improvement. *Journal of Near Infrared Spectroscopy* **12**; 347-358.

Thennadil, S. & Martin, E. 2005. Empirical preprocessing methods and their impact on NIR calibrations: a simulation study. *Journal of Chemometrics* **19**; 77-89.

Tran, C. 2003. Infrared multispectral imaging: principles and instrumentation. *Applied Spectroscopy Reviews* **38**; 133-153.

Trygg, J. 2003. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *Journal of Chemometrics* **17**; 53-64.

Trygg, J. & Wold, S. 2002. Orthogonal projections to latent structures (O-PLS) *Journal of Chemometrics* **16**; 119-128.

Van den Broek, W., Derks, E., Van de Ven, E., Wienke, D., Geladi, P., & Buydens, L. 1996. Plastic identification by remote sensing spectroscopic NIR imaging using kernel partial least squares (KPLS). *Chemometrics and Intelligent Laboratory Systems* **35**; 187-197.

Vogt, F., Banerji, S., & Booksh, K. 2004. Utilizing three-dimensional wavelet transforms for accelerated evaluation of hyperspectral image cubes. *Journal of Chemometrics* **18**; 350-362.

Wang, Y., Veltkamp, D., & Kowalski, B. 1991. Multivariate instrument standardization. *Analytical Chemistry* **63**; 2750-2758.

Weidner, V. & Hsia, J. 1981. Reflection Properties of Pressed Polytetrafluoroethylene Powder. *Journal Optical Society of America* **71**; 856-861.

Westerhuis, J., de Jong, S., Smilde, A. 2001. Direct orthogonal signal correction *Chemometrics and Intelligent Laboratory Systems* **56**; 13-25.

Winson, M., Goodacre, R., Timmins, E., Jones, A., Alsberg, B., Woodward, A., Rowland, J., & Kell, D. 1997. Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR based approach to rapid screening for metabolite overproduction. *Analytica Chimica Acta* **348**; 273-282.

Wold, S., Antti, H., Lindgren, F., & Öhman, J. 1998. Orthogonal signal correction of near-infrared spectra *Chemometrics and Intelligent Laboratory Systems* **44**; 175-185.

Wold, S., Esbensen, K. & Geladi, P. 1987. Principal Component Analysis C*hemometrics and Intelligent Laboratory Systems* **2**; 37-52.

WSC-3. 2004. WSC-3 Proceedings. Nova Science (Ed. A. Pomerantsev) http://rcs.chph.ras.ru/WSC3/book.htm (Accessed July 15, 2006).

*XenICs NV*, Near infrared 2D imaging camera product sheet, Leuven, Belgium. http://www.xenics.com/Products/infrared-imaging.php (Accessed July 15, 2006).

Yu, H. & MacGregor, J. 2003. Multivariate image analysis and regression for prediction of coating content and distribution in the production of snack foods. *Chemometrics and Intelligent Laboratory Systems* **67**; 125-144.

Yu, H. & MacGregor, J. 2004. Monitoring flames in an industrial boiler using multivariate image analysis. *Journal for the American Institute of Chemical Engineers* **50**; 1474-148.

Zhang, L., Henson, M., & Sekulic, S. 2005. Multivariate data analysis for Raman imaging of a model pharmaceutical tablet. *Analytica Chimica Acta* **545**; 262-278.

# I

JOURNAL OF CHEMOMETRICS
*J. Chemometrics* 2005; **19**: 355–363
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.938

# Hyperspectral NIR image regression part I: Calibration and correction

**James Burger* and Paul Geladi**

Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, SLU Röbäcksdalen, P.O. Box 4097, SE 90403, Umeå, Sweden

Received 1 January 2005; Revised 27 February 2005; Accepted 24 July 2005

**Hyperspectral imaging instruments produce large amounts of raw data. These raw data in A/D converter counts have a number of errors that can be corrected by calibration. The use of multiple Spectralon calibration standards is shown to correct for both spectral and spatial variations. Optimal results are achieved using a two-step calibration and correction process. A series of full field of view or external calibration standards is used to transform raw data counts to reflectance values. A grayscale series of internal standards embedded within each hyperspectral image is used to compensate for instrument instability. Second-order regression models based on these multiple standards provide maximum accuracy. The external standards allow for standardization within a hyperspectral image. The internal standards permit instrument standardization or calibration transfer between hyperspectral images. Copyright © 2006 John Wiley & Sons, Ltd.**

**KEYWORDS:** hyperspectral images; multivariate image analysis; instrument standardization; calibration transfer; internal standards; reflectance calibration; multivariate calibration

## 1. INTRODUCTION

Laboratory cameras are nowadays capable of producing high-resolution digital images using ultraviolet, visible, near infrared, infrared, and Raman spectral information [1–4]. Individual images measured with red, green, and blue visible light filters are routinely combined to create 'color' digital images. Sets of images measured at different wavelength bands can be stacked or combined to form hyperspectral images or hypercubes. Using more advanced hardware and software wavelength filtering techniques in combination with broadband imaging detectors, it is now possible to measure tens or hundreds of individual grayscale images or channels with much greater spectral resolution [5–7]. The resulting hyperspectral images can be thought of effectively as collections of tens of thousands of spectra with additional two-dimensional spatial information. Such hyperspectral images have a high potential for describing concentration and property distributions in inhomogeneous solids, semisolids, and liquids. Spectroscopic multivariate calibration modeling techniques can be extended to hyperspectral images, permitting qualitative and quantitative analysis. But accurate results require accurate spectra. The objective of this paper is the presentation and discussion of the use of proper external and internal calibration standards in hyperspectral images. Results are based on recalibration of known reference standards. Previous work discussed the use of external standards only [8].

Much work has been published regarding the standardization of spectroscopic instruments and the transfer of multivariate calibration models [9–14]. Three situations can occur which invalidate calibration models: (1) physical or chemical changes in samples such as texture or particle size, (2) instrument changes due to inherent instabilities, aging parts or the use of multiple instruments, or (3) environment changes such as temperature or humidity [13]. The resulting spectral changes may be observed in either the horizontal axis as wavelength shifts or peak broadening, in the vertical axis as baseline offsets or absorbance differences.

Calibration model transfer of hyperspectral images is subject to all the error contributions of conventional one-dimensional spectroscopy (noise, drift, non-linear response of detectors, wavelength-dependent errors) as well as the two-dimensional or spatial error components associated with camera devices and illumination (readout errors, inconsistent detector responses, quantization errors, and non-uniform lighting).

Spatial and spectral standards are very common in photography and television/video work. Gray cards, grayscale wedges, color scales, and other test images can be used to test correctness of the obtained images for different illuminations, film types, or camera settings. Other test standards for resolution, geometry etc. are also available. But similar standard materials are not easily found for the NIR region. One source is a calibration surface made of Spectralon [15]

*Correspondence to: J. Burger, Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, SLU Röbäcksdalen, P.O. Box 4097, SE 90403, Umeå, Sweden.
E-mail: jim.burger@btk.slu.se

available with different levels of reflectance with calibration values ranging from 250 to 2500 nm. These UV-VIS-NIR reflectance standards are intended for checking and calibrating spectrometers with a typical reflection geometry, but in this paper they are tested as image reflectance standards.

The advent of the increased dimensionality of hyperspectral images has provoked a parallel development in chemometrics tools. PCA has been extended to multivariate image analysis (MIA) [16,17]. Likewise, regression techniques such as OLS, PCR, and PLS have been incorporated into MIR [18–20]. In MIA a common technique is to use PCA of images reorganized to matrices. This allows the use of loadings to interpret the importance of the wavelengths and the use of score images and score plots in conjunction with segmentation to detect concentrations or property regions or gradients. Alternatives to PCA such as factor analysis or curve resolution can also be used. In MIR, the hypercube is regressed against an image of external information, for example concentrations in a calibration step. The goal is to be able to use a hypercube to predict concentrations once the regression model is established.

The accuracy of any modeling or prediction method is limited by the quality of the data used. MIA and MIR can be applied to the raw data given in A/D converter counts, but these counts are not in SI or IUPAC spectroscopic units (reflectance or absorbance), nor are they very reproducible. This results in an under usage of the full potential of MIA and MIR. It is necessary to make the hypercubes consistent both internally (within a hypercube) and also externally (between hypercubes).

Preprocessing steps must be taken to clean or prepare the data for further analysis. Spectral data are routinely transformed to reflectance using a simple two-point transfer function based on dark current and maximum reflectance value measurements. The use of additional calibration standards can improve measurement accuracy and permits linear and non-linear corrections to be made. The current work demonstrates the use of both external and internal calibration standards. External standards are used to correct for pixel-to-pixel variances due to camera inconsistencies and variation in sample illumination. Internal standards are used to compensate for signal drift over time due to changes in power or temperature effects.

## 2. THEORETICAL BACKGROUND

### 2.1. Reflectance transforms

Spectral data collected from a CCD device represent detector signal intensity counts not actual reflectance values. Figure 1 shows the typical raw data spectra for a series of NIR materials, which should exhibit flat and smooth uniform spectra. The apparent structure and variation in signal intensity of these spectra is due to the wavelength dependencies of both the sample illumination source and the sensitivity of the detector used. It is generally more useful to correct or transform raw signal data into reflectance or absorbance units by comparing with spectra of standard reference materials. The usual transformation to reflectance values is obtained by correcting sample spectra (C) for detector dark current (Dk) and dividing by a similarly corrected total reflectance spectrum ($Ref_{100}$) [2,6].

$$R = (C - Dk)(Ref_{100} - Dk)^{-1} \quad (1)$$

Equation (1) transforms instrument measurement values in A/D counts ( C, Dk, and $Ref_{100}$ ) into unitless reflectance values. This transform is often termed a one-point calibration since the transformation is based on a single reference standard, $Ref_{100}$. Mathematically, however, this is a two-point

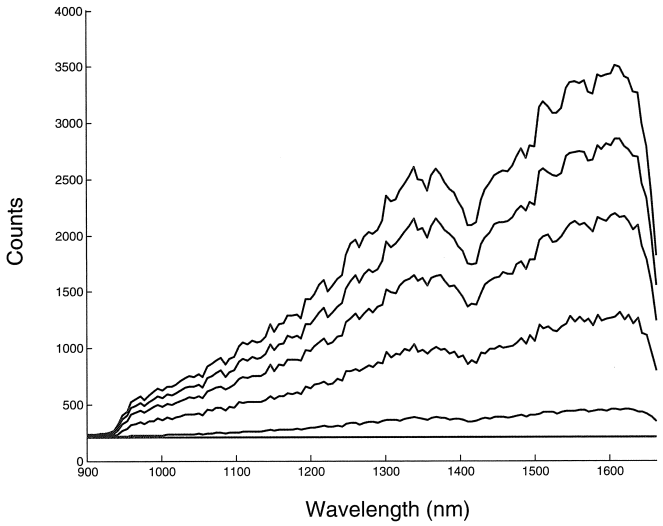


**Figure 1.** Average spectra (raw counts) of 99 (top curve), 75, 50, 25, 2% Spectralon and dark current (baseline).

linear approximation model with measured endpoint values at 0 and 100% reflectance and for the purpose of this discussion will be termed a *simple* model. Equation (1) can be rearranged and written in a linear bias offset and slope form:

$$R = -\mathrm{Dk}(\mathrm{Ref}_{100} - \mathrm{Dk})^{-1} + (\mathrm{Ref}_{100} - \mathrm{Dk})^{-1}\mathrm{C} \qquad (2)$$

Or in general terms:

$$R = b_0 + b_1\mathrm{C} \qquad (3)$$

This *simple* transformation model requires only two calibration spectra: one of the total reflectance standard and a dark current spectrum measured by blocking the camera lens. Such a model covers the complete range of reflectance values but gives no indication of system non-linearity.

Optimal transform models are obtained when more than two standards can be measured which span the reflectance or absorbance space of the expected samples. When *true* calibration spectra are available for a collection of standards, an alternative modeling approach is available: fitting least squares regression models to the true versus measured values for the set of all standards. In this paper we present two such regressions: a first order *linear* model with coefficients as in Equation (3) and a second-order quadratic or *quad* model, Equation (4).

$$R = b_0 + b_1\mathrm{C} + b_2\mathrm{C}^2 \qquad (4)$$

Although the form of the equations for the *simple* and *linear* transformation models is the same, the $\boldsymbol{b}_\mathbf{i}$ coefficients are computed differently. They are based on a single standard and dark measurement for the *simple* model, and more than two standards for the *linear* model. Since the measured response of the system is highly wavelength dependent, individual *simple*, *linear*, and *quad* models must be computed at each wavelength. For the transformation of multiple wavelength spectra, the scalar model coefficients $b_i$ in Equations (3) and (4) are therefore reported as vectors $\boldsymbol{b}_\mathbf{i}$ as a function of wavelength. It is also important to note that no dark subtraction is performed for the linear and quad regression models. The dark offset is included as part of the $\boldsymbol{b}_\mathbf{0}$ model constant.

## 2.2. Spatial modeling—external standards

Hyperspectral images add spatial complexity to the spectral transformation problem. Regardless of whether the hypercubes are acquired from a single detector, or a one- or two-dimensional array of detectors, spatial changes in sample illumination or differences in multiple detector sensitivities require that the wavelength-dependent reflectance transform models must be computed for each spatial or pixel location. Two approaches will be considered, which we term *global calibration* and *pixelwise calibration*. Both techniques are spatial extensions to the one-dimensional reflectance transform problem, and necessitate accumulation of complete hypercubes from a series of calibrated reference materials. These hypercubes are measured independently of subsequent sample images and are therefore termed *external standards*.

The *global calibration model* assumes only wavelength dependencies and essentially ignores all spatial dependencies, assuming uniform sensitivities throughout each wavelength channel of the hypercube. This approach is based on the median spectrum of each of the individual calibration standard hypercubes. Figure 2 shows the computation steps necessary to create each model. For each of the calibration standards, a single median spectrum is first computed from the collection of spectra of every pixel in the respective hypercube. This set of median spectra is then used to build the global calibration model at each wavelength. As was described previously, the *simple* transform model is based only on total reflectance and dark hypercubes, while the *linear* and *quad* regression models are based on sets of calibrated reference standards. Future hypercubes are then transformed to reflectance, pixel by pixel, using the wavelength-dependent global model $\boldsymbol{b}_\mathbf{i}$ vectors.

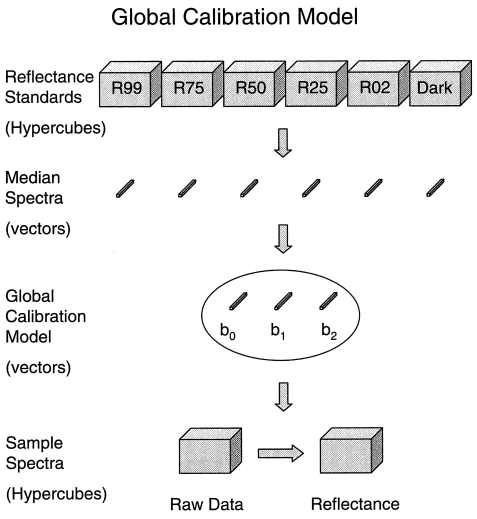


**Figure 2.** The median spectrum from each of the external standard hypercubes is used to compute a global calibration model one wavelength at a time. The resulting model vectors (individual $b_0$ and $b_1$ of Equation (3) and $b_0$, $b_1$, and $b_2$ of Equation (4) at each wavelength) are then used to transform sample hypercubes from instrument counts (raw data) to reflectance at each wavelength.

Independent reflectance transform models can also be created for each individual pixel in the hypercube. This compensates for illumination and detector inhomogeneities. Rather than using the median spectra of the global image, position-specific transform models are computed one pixel at a time from the same pixel location in each of the external standard hypercube images. Figure 3 represents a flowchart for the computation steps of this model process. The complete set of individual regression coefficient vectors computed at each pixel effectively form hypercubes of regression coefficients $B_i$, in contrast to the global regression vectors $\boldsymbol{b}_\mathbf{i}$ in Figure 2. This spatial specific reflectance transform model is termed a *pixelwise calibration model*.

## 2.3. Internal standards

The use of external calibration standards as discussed above permits the creation of spatial and wavelength-specific

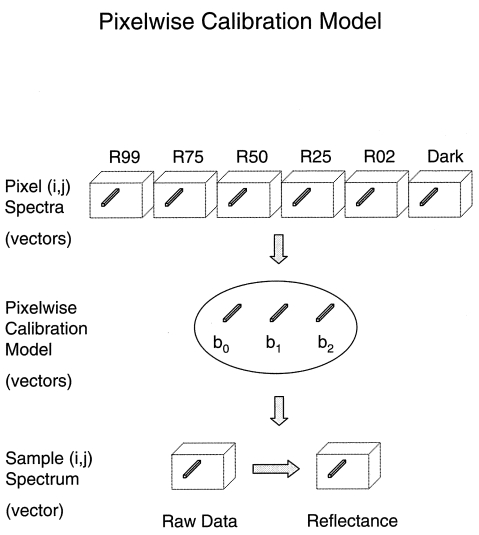


**Figure 3.** Each of the individual (*i,j*) spectra from the external standard hypercubes are used to compute individual pixel calibration model vectors ($b_0$ and $b_1$ of Equation (3) or $b_0$, $b_1$, and $b_2$ of Equation (4) at each wavelength). These pixel (*i,j*) specific model vectors are then used to transform the respective (*i,j*) raw data sample spectra from instrument counts to reflectance.

reflectance transform models. Such models compensate for effects due to differing sensor response sensitivities, sensor non-linearity, or non-uniform lighting. But other factors influence the image quality as well. Thermal drift of the detector and filters, age or temperature effects on the lamps, or power supply voltage fluctuations can all contribute to degradation of image quality. These factors and others have been addressed collectively in spectroscopy as instrument standardization or calibration transfer issues.

One of the advantages of hyperspectral images is the immense amount of data available. This enables large sample statistics to be computed. It also permits the opportunity to target some of the data space specifically for calibration or diagnostic purposes. Adding a small mosaic of internal standard materials, which fills only 5% of the field of view of an $320 \times 256$ pixel image, creates over 4000 spectra in every hypercube to perform data validation tests.

If the internal standards are calibrated and have known spectra, then correction models may be computed to realign the measured spectra. Corrections in both wavelength or peak position and reflectance or absorbance intensities can be determined. An alternative approach used in this paper is to simply declare one sample hypercube 'master' and adjust other 'slave' sample hypercubes to match. Normalization of measured intensities is performed by using either a single-point stretch (Equation 5) or a multiple-point linear (Equation 6) or quadratic (Equation 7) regression-based adjustment, depending on the number of internal standards available:

$$R_c = \alpha R \tag{5}$$

$$R_c = \alpha_0 + \alpha_1 R \tag{6}$$

$$R_c = \alpha_0 + \alpha_1 R + \alpha_2 R^2 \tag{7}$$

Here $R_c$ is the corrected reflectance value. The $\alpha_i$ constants are computed from the measured master versus slave values. These internal standard-based corrections are global corrections applied to the entire image. That is, a single correction model based on internal standard ROIs is used to correct the reflectance values of all pixels in the image for a single wavelength channel. These correction models are of course wavelength channel dependent, and therefore $\alpha_i$ vectors must be computed to correct each complete hypercube. This is equivalent to piecewise direct standardization (PDS) with a window width of one [11]. It is assumed that spatial variation dependencies have been previously removed by the external standard-based reflectance transform. Consequently, positional registration of the internal standards from hypercube to hypercube is not necessary.

Figure 4 indicates the two primary steps necessary for these adjustments. A sample raw data hypercube C is first

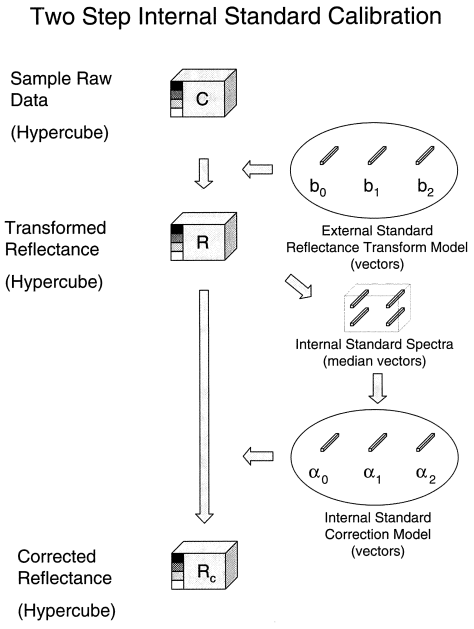


**Figure 4.** Sample hypercubes C are first transformed to reflectance *R* using the pixelwise wavelength-specific regression coefficient vectors from the external standards calibration model. Median spectra from each of the internal standard regions are then used to compute a second wavelength-specific correction model ($\alpha_i$ of Equations (5–7) at each wavelength), which is then applied to all spectra of the current reflectance hypercube to compute the corrected reflectance hypercube $R_c$.

transformed to reflectance $R$ using an external standard-based model. Spatial regions of interest are then identified for each internal standard and median spectra computed for each internal standard. These median spectra are then compared with either known calibration spectra or spectra from a master hypercube to create a second correction model, which is then applied globally to each spectrum of the hypercube, resulting in the corrected reflectance hypercube $R_c$. This correction model is specific to only one hypercube and must be recomputed for each subsequent hypercube. For the purpose of discussion, as with the reflectance transform models, the internal standard-based correction models will be termed *simple*, *linear* or *quad* to represent one-point, or multi-point first- or second-order regression calibration models.

## 3. EXPERIMENTAL/MATERIALS AND METHODS

### 3.1. Hyperspectral images

NIR hyperspectral images were collected using a Spectral Dimensions matrix NIR instrument. This instrument uses a liquid crystal tunable filter (LCTF) bandpass filter in combination with an InGaAs diode array detector [21]. Images are digitized with 12 bit resolution. Each image was measured with a 32 ms integration time, repeated 10 times, and averaged. Individual images were acquired between 900 and 1662 nm at 6 nm intervals. This set of individual images when stacked, created a $320 \times 256 \times 128$ hyperspectral image, or hypercube.

Four quartz halogen lamps were positioned to provide as uniform as possible illumination of the image field of view ($62 \times 50$ mm). The halogen lamps produce a significant amount of heat and their use at maximum power is only possible for inorganic materials. For this study, the lamp power was initially set at 60%, and then sequentially reduced for some images. Individually calibrated tiles of Spectralon standard NIR reflectance material [15] were used for image calibration and test samples. Reference spectra were computed as parabolic interpolations from certified tables provided by Labsphere (Sutton, NH) [8]. The Spectralon material for these samples (99, 75, 50, 25, and 2% reflectance) is created by adding an increasing amount of carbon black to a white Teflon base material, which at high magnification appears inhomogeneous and textured. To avoid imaging these non-uniformities, the external standard samples were positioned on a rotating bearing and manually rotated during image acquisition. Samples were also positioned 5 cm above the normal camera focal plane to purposely blur any surface defects.

### 3.2. Experiment 1: external standards

To test the effectiveness of the different reflectance transform functions a set of *external standard* hyperspectral images was measured where the large spinning spectralon tile filled the entire camera field of view. (Images R99, R75, R50, R25, and R02) This provided hypercubes of uniform material at every pixel at all wavelengths. An additional uniform dark hypercube was measured by simply blocking the entrance to the camera. Replicate images were measured to allow the construction of different calibration and test datasets.

### 3.3. Experiment 2: internal standards

Effects of lamp aging and instability were simulated by a drastic reduction of the lamp power. In order to study corrections for this phenomenon a series of *internal standard* images (a mosaic of multiple Spectralon samples) was measured while systematically varying the lamp power level. Four of the Spectralon tiles (25, 50, 75, and 99% reflectance) were arranged to cover the four corners of the field of view. Ten stationary hypercube images of this Spectralon mosaic were then measured while decreasing the lamp power to 60, 59, 58, 57, 56, 55, 50, 45, 40, and 35% power. Spatial surface irregularities in these stationary images were compensated for by computing median spectra of large internal standard region of interest (ROI) areas.

### 3.4. Data analysis

It is common for InGaAs diode array detectors to have a few elements that respond with unusually high or low values commonly referred to as *dead pixels*. These pixels have high leverage and can lead to erroneous results in calibration models. Rather than compute global average values, median spectra were computed to eliminate the high leverage effects of the outlier dead pixels. Sorting the intensities measured within an ROI by value effectively places the dead pixel measurements at the endpoints of the measurement range, minimizing the impact on median computation. Alternatively, these outlier pixels were detected by using simple univariate thresholding techniques and eliminated from any computational work. All references to average spectra in this paper assume that such outliers have been detected and removed before averaging. No other data pretreatment was performed. All computations were performed using MATLAB version 6.1 [22].

## 4. RESULTS AND DISCUSSION

### 4.1. External calibration

Global and pixelwise reflectance transform models were computed from the calibration hypercubes of the Spectralon standard reference materials. The R99 and dark current hypercubes were used to create the simple model. An additional $\alpha$ correction term was added to compensate for the fact that a 99% standard was used to represent total reflectance. This is effectively a combination of Equations (3) and (5):

$$R_c = \alpha(b_0 + b_1 C) \tag{8}$$

The correction term $\alpha$ is wavelength dependent and is equal to the reported reflectance value for the R99 standard reference material and applies to the *simple* model only.

The R99, R75, R50, R25, and R02 hypercubes were used to create linear and quad regression models. Every spectrum in a replicate R50 test hypercube was then individually transformed to reflectance using the global-based models. The median of these reflectance spectra is presented in Figure 5. In contrast to the highly irregular raw counts spectra of Figure 1, fairly flat reflectance spectra are obtained as expected. The deviations in the 900–950 nm range are attributed to the very low signal to noise ratio of this region. It can be clearly seen, however, that the reflectance transform models based on regression models of multiple standards

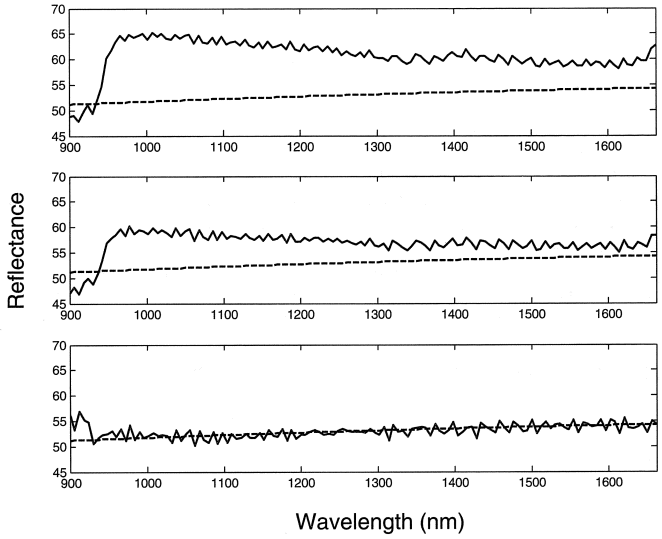


**Figure 5.** Average predicted reflectance spectra of 50% Spectralon based on global image models of calibration standards. Simple (top), linear (middle), and quad (bottom) models versus true reflectance spectra (- - -).

improve the predicted results. The spectrum based on the second-order quad model was nearly indistinguishable from the expected true spectrum. Similar results were obtained with replicate images of the other calibration standards.

Figure 6 shows a comparison of the results for the global- versus pixelwise-based transforms of a replicate R75 hypercube image. The top plot is based on the quad models of the global image median spectra; the bottom plot is based on the

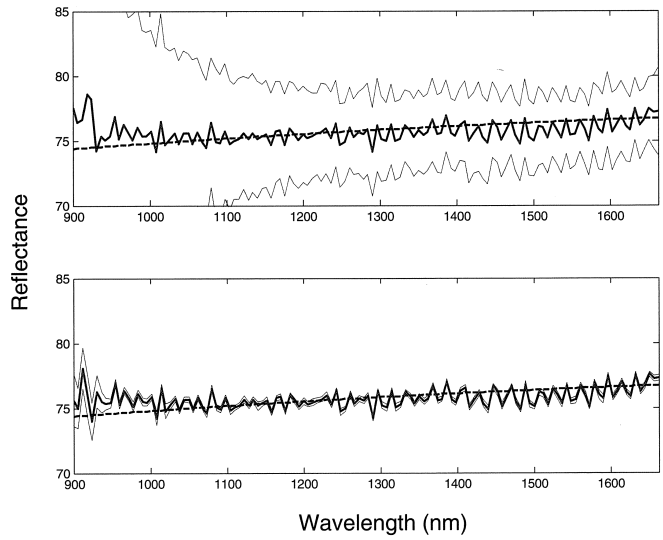


**Figure 6.** Average predicted spectra of 75% Spectralon (dark line) including standard deviation bands (lighter lines) based on global image quad models (top) versus individual pixelwise quad models (bottom) of calibration standards versus true reflectance spectra (- - -).

**Table I.** External standard calibration. Average measurement errors (percent reflectance) for the 75% Spectralon standard, 1100–1600 nm

| | Model | Bias | Standard deviation |
|---|---|---|---|
| Global | Simple | 4.65 | 2.63 |
| | Linear | 0.91 | 2.78 |
| | Quadratic | −0.34 | 3.24 |
| Pixelwise | Simple | 4.75 | 0.21 |
| | Linear | 1.00 | 0.19 |
| | Quadratic | −0.26 | 0.23 |

pixelwise quad models computed at each individual pixel location. There is a slight improvement in the lower wavelength range, but average spectra results for the two modeling techniques are quite similar. This is because the two single spectra plotted represent the average of thousands of spectra measured from a uniform material, with fairly uniform lighting. The standard deviation spectra show the benefit of the pixelwise modeling: the standard deviations have been significantly reduced. Any deviations between individual sensors of the diode array detector, or differences in lighting effects, have been reduced by the individual pixel modeling. A more typical hyperspectral image containing multiple objects with differing spectral components would clearly benefit from this pixel-specific modeling advantage.

Table I summarizes the average results for predicting the 75% reflectance material between 1100 and 1600 nm. A comparison of the global and pixelwise models indicates nearly identical bias values; however, the standard deviations are reduced by a factor of 10 for the pixelwise models. For non-uniform sample illumination this difference would be even more apparent. Simple calibration models based on only dark current and a single external standard were the least effective. The linear and quad regression model transforms based on multiple external standards performed significantly better. Maximum accuracy for the reflectance transform model was achieved using multiple standards (without a need for a dark current measurement) and computing second-order regression models for each pixel at each wavelength.

## 4.2. Internal calibration

The inclusion of internal standards was examined using a series of 10 hypercubes where the light illumination intensity was varied by purposely changing the lamp power. The series of reduced power hyperspectral images was first transformed to reflectance using the pixelwise-based quad models based on the previously measured external standard hypercubes at 60% power. Four ROIs were then identified and the median value of each region at each wavelength determined. Again median spectra were computed to minimize leverage contributions of dead pixels and sample inhomogeneity effects of the stationary materials imaged. The data displayed in Figure 7 represent a single-wavelength channel (centered at 1602 nm) but plots at other wavelengths showed similarly shaped response curves. The response signals increase non-linearly with an increase in lamp power (filament temperature) as indicated in the measured reflectance values.

Figure 8 shows the median spectra of the R75 reflectance standard ROI for the eight lamp power settings between 60 and 45%. Note the change in vertical axis scaling for viewing purposes. The top plot shows that decreasing the lamp power reduced the uncorrected reflectance spectrum from

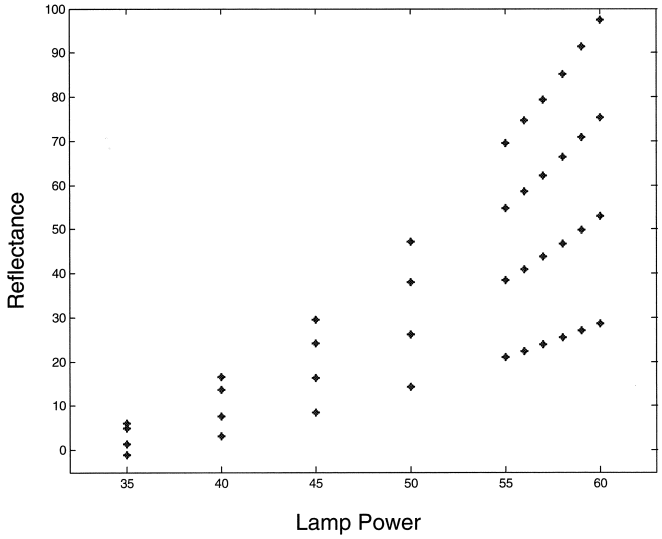


**Figure 7.** Median reflectance values (1602 nm) versus lamp power for 99 (top), 75, 50, and 25% Spectralon.

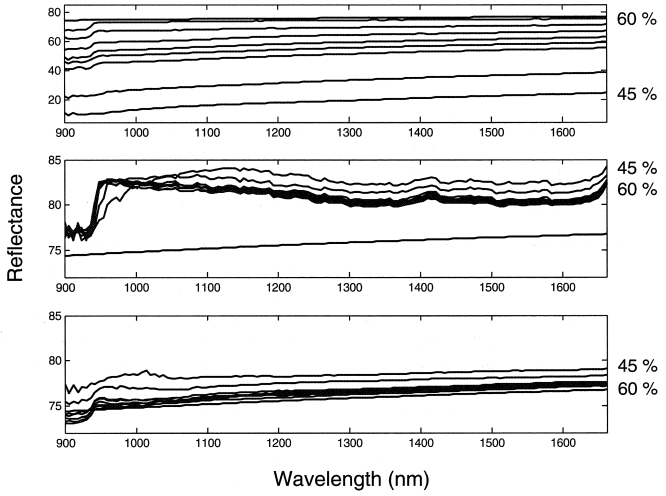


**Figure 8.** Average reflectance spectra of 75% Spectralon. Uncorrected (top) versus single point (middle), and quad (bottom) corrections based on internal calibration standards. Lamp power sequentially reduced from 60 to 45%. Note change in vertical axis scaling.

the expected 75% to between 10% and 20% reflectance values. The middle plot shows the corrected results using a single internal standard. In this case $\alpha$ scaling factors were computed by scaling the measured median spectra of the R99 internal standard back to the expected value (Equation 5). This simple approach resulted in corrected R75 spectra with only 5%–10% deviations from expected values.

The bottom plot in Figure 8 shows the effect of a quad fit to the median spectra of the R99, R50, and R25 standards (Equation 7). Reducing the lamp power from 60% to 55% reduced the uncorrected spectra from the expected 75% to roughly 40% reflectance (Figure 8 top). But adding the second quad-based internal standard correction resulted in a corrected spectra with only 1–2% error (Figure 8 bottom). Even the 45% lamp power spectra were corrected to within 4% error.

Table II summarizes the results for adjustments to the 75% reflectance standard based on models of the other internal standards. Reducing the lamp power from 60 to 45% is a drastic change, yet even the one-point simple adjustment corrects the bias to within 4–7%. In all cases the bias and standard deviation values increase with a reduction in power, as would be expected from heteroscedastic noise in the system. The best results were obtained with the second-order quadratic correction model. In this example, the correction models were based on only three internal standards, used to predict the values of the 4th standard. Even better results would be expected from creating second-order correction models based on all four or more internal standards.

**Table II.** Internal standard calibration. Average measurement errors (percent reflectance) for the 75% Spectralon standard, 1100–1600 nm; bias (standard deviation)

| Power | Uncorrected | Simple | Linear | Quadratic |
|---|---|---|---|---|
| 60 | −1.55(0.17) | 4.45(0.50) | 0.63(0.07) | 0.38(0.04) |
| 59 | −6.32(0.69) | 4.54(0.51) | 0.78(0.09) | 0.59(0.06) |
| 58 | −11.47(1.26) | 4.70(0.53) | 0.88(0.10) | 0.69(0.08) |
| 57 | −16.27(1.79) | 4.84(0.54) | 0.97(0.11) | 0.79(0.09) |
| 56 | −19.97(2.20) | 4.91(0.55) | 1.11(0.13) | 0.98(0.11) |
| 55 | −24.08(2.65) | 5.11(0.57) | 1.17(0.13) | 0.99(0.11) |
| 50 | −41.79(4.59) | 5.90(0.65) | 1.99(0.22) | 1.80(0.20) |
| 45 | −55.47(6.09) | 6.87(0.76) | 2.99(0.33) | 2.58(0.28) |

## 5. CONCLUSIONS

Hyperspectral images contain a rich mixture of both spectral and spatial information. For effective use of these images to identify chemical components and to determine their concentration distributions, accurate calibration steps must be performed to transform measured detector responses into reflectance or absorbance values. This requires calibration by the use of standards such as Spectralon NIR standard reflectance material. Problems with surface defects and spatial variations in material composition can be easily corrected. Sample variations can be corrected by either spinning the reference material during image acquisition, or averaging the hundreds or thousands of spectra measured within an internal standard image region. The use of this material as both external and internal standards permits the uniform and accurate transformation of signal counts to reflectance at all pixel locations and at all wavelengths of the hyperspectral image.

Three aspects of hyperspectral image calibration were considered: (1) data-subset selection (global vs. pixelwise modeling), (2) data model type (simple two point linear fit,

or linear or quadratic regression), and (3) type of standards selected (external or internal). Results from these options are summarized in Tables I and II.

The tables show that data subset selection was optimized using the pixelwise models, allowing for corrections of both spatial and wavelength-dependent variations due to non-uniform lighting problems or variations in individual diode sensor sensitivity. Data model type was optimized by the use of multiple external reflectance standards coupled with pixelwise second-order regression models. This combination is recommended for the accurate transformation of hyperspectral images from instrument counts to reflectance values.

Use of internal standards incorporated into the image allowed for detection and correction of time-dependent variations due to temperature- or power-induced changes to the overall image. As with the external standard calibrations, corrections based on only the 99% standard were least effective. Optimum results were obtained using an array of internal standards to compute a first- or second-order regression-based correction model. (Table II) Any measurement variations within the internal standard material were simply removed by computing median spectra of each internal standard ROI. This technique also eliminated any need for accurate registration or placement of the internal standard sample locations from image to image.

The primary goal of this study was the testing of external and internal standards for calibrating and correcting NIR hyperspectral images. The Spectralon standards used in this study all had fairly flat reflectance spectra throughout the NIR region and were selected to span reflectance values between 99 and 2%. The same material was used for both calibration and test samples. But there is a clear need for the development of other secondary NIR standards. These standards should similarly span high and low reflectance values, and be easily calibrated with external reference spectra. Any sharp spectral features would enable testing of wavelength accuracy. This study focused on the use of spectral standards for use in correcting hyperspectral image intensities. The design of image standards for spatial or geometry calibration is also needed. Further development of such NIR hyperspectral image standards is ongoing.

An application of the use of external and internal standard-based corrections to hyperspectral NIR image regression will be presented in Part II.

## Acknowledgments

The authors acknowledge financial support from NIRCE, an EU Unizon-Kvarken project. We also acknowledge Torbjörn Lestander, Swedish University of Agricultural Sciences and Robert Cogdill, Iowa State University for related discussions.

Kempestiftelserna grant SMK-2062 provided funds for the equipment.

## REFERENCES

1. Wolfe W. *Introduction to Imaging Spectrometers*. SPIE Optical Engineering Press: Bellingham, Washington, USA, 1997.
2. Polder G, vander Heijden G, Keizer L, Young I. Calibration and characterisation of imaging spectrographs. *J. Near Infrared Spectrosc.* 2003; **11**: 193–210.
3. Martinsen P, Schaare P, Andrews M. A versatile near infrared imaging spectrometer. *J. Near Infrared Spectrosc.* 1999; **7**: 17–25.
4. Brianco A, Serafino G, Spoeck G. *An introduction to spectral imaging*. http://www.ctr.at/hypspbvdoc/IntSpImg.pdf. 2005.
5. Hoelter T, Barton J. Extended short wavelength spectral response from InGaAs focal plane arrays. *Proc. SPIE* 2003; **5074**: 481–490.
6. Gat N. Imaging spectroscopy using tunable filters: a review. *Proc. SPIE* 2000; **4056**: 50–64.
7. Tran CD. Infrared multispectral imaging: principles and instrumentation. *Appl. Spectrosc. Rev.* 2003; **38**: 133–153.
8. Geladi P, Burger J, Lestander T. Hyperspectral imaging: calibration problems and solutions. *Chemometr. Intell. Lab. Syst.* 2004; **72**: 209–217.
9. Wang Y, Veltkamp D, Kowalski B. Multivariate instrument standardization. *Anal. Chem.* 1991; **63**: 2750–2756.
10. Wang Y, Kowalski B. Calibration transfer and measurement stability of near-infrared spectrometers. *Appl. Spectrosc.* 1992; **46**: 764–771.
11. Wang Y, Lysaght M, Kowalski B. Improvement of multivariate calibration through instrument standardization. *Anal. Chem.* 1992; **64**: 562–564.
12. Fearn T. Standardisation and calibration transfer for near infrared instruments: a review. *J. Near Infrared Spectrosc.* 2001; **9**: 229–244.
13. Feudale R, Woody N, Tan H, Myles A, Brown S, Ferre J. Transfer of multivariate calibration models: a review. *Chemometr. Intell. Lab. Syst.* 2002; **64**: 181–192.
14. Geladi P. Some recent trends in the calibration literature. *Chemometr. Intell. Lab. Syst.* 2002; **60**: 211–224.
15. Spectralon Diffuse Reflectance Targets, Labsphere, Inc., PO Box 70, 231 Shaker Street North Sutton, NH 03260, USA.
16. Geladi P, Wold S, Esbensen K. Image analysis and chemical information in images. *Anal. Chim. Acta.* 1986; **191**: 473–480.
17. Esbensen K, Geladi P. Strategy of multivariate image analysis. *Chemometr. Intell. Lab. Syst.* 1989; **7**: 67–86.
18. Geladi P, Esbensen K. Regression on multivariate images: principal component regression for modeling, prediction and visual diagnostic tools. *J. Chemometr.* 1991; **5**: 97–111.
19. Esbensen K, Geladi P, Grahn H. Strategies for multivariate image regression. *Chemometr. Intell. Lab. Syst.* 1992; **14**: 357–374.
20. Lied T, Geladi P, Esbensen K. Multivariate image regression (MIR): implementation of image PLSR—first forays. *J. Chemometr.* 2000; **14**: 585–598.
21. Spectral Dimensions, Inc., 3416 Olandwood Court, Suite 210, Olney, MD 20832, USA.
22. The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098, USA.

# II

# Hyperspectral NIR image regression part II: Dataset preprocessing diagnostics

**James Burger* and Paul Geladi**

Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, PO Box 4097, SE 90403 Umeå, Sweden

When known reference values such as concentrations are available, the spectra from near infrared hyperspectral images can be used for building regression models. The sets of spectra must be corrected for errors, transformed to reflectance or absorbance values, and trimmed of bad pixel outliers in order to build robust models and minimize prediction errors. Calibration models can be computed from small (< 100) sets of spectra, where each spectrum summarizes an individual image or spatial region of interest (ROI), and used to predict large (> 20,000) test sets of spectra. When the distributions of these large populations of predicted values are viewed as histograms they provide mean sample concentrations (peak centers) as well as uniformity (peak widths) and purity (peak shape) information. The same predicted values can also be viewed as concentration maps or images adding spatial information to the uniformity or purity presentations. Estimates of large population statistics enable a new metric for determining the optimal number of model components, based on a combination of global bias and pooled standard deviation values computed from multiple test images or ROIs. Two example datasets are presented: an artificial mixture design of three chemicals with distinct near infrared spectra and samples of different cheeses. In some cases it was found that baseline correction by taking first derivatives gave more useful prediction results by reducing optical problems. Other data pretreatments resulted in negligible changes in prediction errors, overshadowed by the variance associated with sample preparation or presentation and other physical phenomena.

**KEYWORDS**: Hyperspectral images, multivariate image regression, partial least squares modeling, spectral preprocessing, regression model validation

## 1. INTRODUCTION

In a previous article [1] it was explained how to convert hyperspectral images usually collected as A/D converter counts into corrected diffuse reflectance values. For this correction purpose, external (between-image) and internal (within-image) Spectralon [2] UV-Vis-NIR standard reflectance materials were used together with two-point, linear and nonlinear correction equations. One objective of these corrections is to standardize spectra for purposes of visual comparisons and to make spectra more interpretable. Accurate spectra are also essential for creation of robust and reliable discriminant analysis and regression models. Spectra are sometimes normalized or subjected to derivative transforms to correct for baseline offsets and physical effects within the sample [3,4]. In this paper the impact of the internal and external standard based corrections on calibration models and their effect on prediction errors is examined. Different approaches to hyperspectral data preprocessing and their effect on model formation and predictive performance are also explored.

Hyperspectral images offer an abundance of data. Many pixels or spectra are available even in a local scene or spatial region of interest (ROI). This large population of calibration and test objects requires a different look at modeling and diagnostics. Although the number of calibration and test sample spectra is huge, the set of available reference material values is relatively very small, requiring extra care in model validation. An important property of calibration models is the determination of pseudorank, the number of latent variables or components to incorporate in the model [5]. The evolution of prediction bias and prediction variance as the number of model components increases is investigated. The small set of sample reference values necessitates leave-one-out sample cross validation. Regression diagnostics are often based on prediction bias alone, but may include variance as well [6,7]. Because of the large number of pixel samples in the hyperspectral images, prediction results can be studied as histogram distributions and large population statistics estimated. A new metric for the pseudorank determination based on these estimates is described and tested.

By applying class membership masks, partial least squares (PLS) discriminant analysis has been applied to hyperspectral images to classify unknown image pixels [8,9]. Extraction of image features with principle component analysis (PCA) followed by PLS or artificial neural networks (ANN) has been used to obtain global quantitative information [10-12]. To apply PLS directly to a hyperspectral image, individual Y-block reference values are needed at every pixel of all calibration images. If such detailed data is unavailable, then global or mean reference values can be used, if each reference value is properly matched with a mean spectrum from a single image or large ROI representing the known sample. The resulting calibration model can then be used to make predictions at individual pixels of test images. The use of such ROI based models is explored in this paper.

Two examples are used to illustrate how models are built and how they perform. Artificial mixtures of laboratory chemicals were prepared with the concentrations in the

mixtures determined in advance by gravimetric analysis. A second sample set was based on a collection of commercial cheeses, where concentrations of protein, fat and carbohydrate were indicated on the packaging provided by the manufacturer, and alternatively measured in the laboratory (fat, protein). The global or average ingredient concentrations are determined as with other spectroscopic analysis techniques, but since hyperspectral images provide additional spatial information, creation of concentration maps or images is possible. We show that the examination of the population distributions of predictions resulting from hyperspectral image regression provides additional diagnostic information, permitting further characterization of samples and performance of the instrument.

## 2. THEORETICAL BACKGROUND

### 2.1 Nomenclature

The following notation is used in order to facilitate discussion:

$\underline{\mathbf{Z}}$ : a typical hyperspectral image or hypercube (I×J×K variables). This hypercube contains K single channel or gray scale images ($\mathbf{Z}_k$) each containing I rows and J columns of pixels ($z_{ijk}$). The same hypercube can be viewed as I * J spectra ($\mathbf{z}_{ij}$), each containing K variables.

K : the number of wavelength bands in a hypercube, or variables in a spectrum.

L : the number of pixels within one region of interest (ROI) in one image.

P : the total number of hypercube images or ROIs.

Q : the total number of replicate images or ROIs for one sample.

N : the number of calibration objects. N can be I * J if one whole image is used, it may be another number of spectra (N=L if a region of interest is used), it may be the sum of the number of spectra from a number of regions of interest from the same or different images ($N=L_1+L_2+L_3... L_P$) or it may be the number of images (P) or regions used if they are each represented by individual mean or median spectra.

M : the number of test set objects. The same explanation as for N is also valid here. In some cases M >> N in the present paper.

A : the number of latent variables in a PLS model

### 2.2 Regression models and prediction

Hyperspectral image regression involves the building of regression models between hyperspectral images and external data. This means that for every image pixel containing the spectral vector $\mathbf{x}_{ij}$ there must be related external data $y_{ij}$. A spatial mask may sometimes be used to select subset regions $\underline{\mathbf{Z}}_r$ (r stands for region of interest) of the image to formulate optimized local models. However specific spatial information for the external $y_{ij}$ reference values may not be available for individual pixel locations. In this case a single global y value may be used for all L samples in $\underline{\mathbf{Z}}_r$. Alternatively, $\underline{\mathbf{Z}}_r$ may be reduced to a single spectrum or sample $\mathbf{z}_r$ representing the mean or median of $\underline{\mathbf{Z}}_r$. A calibration set can then be constructed from either the $\underline{\mathbf{Z}}_r$ or $\mathbf{z}_r$ obtained from multiple regions from one or more hyperspectral images, and their respectively ordered y values. If **y** and **X** are variable-wise mean-centered, a regression model may be formed:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{f} \qquad (1)$$

**y** : external data for N spectra from one or more hyperspectral images, ordered in a vector

**X** : spectra ordered in the same way as in **y** to form a matrix (N×K)

**b** : a vector of regression coefficients (K×1)

**f** : the residual (N×1)

If the model in Equation 1 is adequate, the regression vector can be used for prediction of new hyperspectral images taken under the same physical conditions:

$$\mathbf{y}_h = \mathbf{X}_t\mathbf{b} \qquad (2)$$

$\mathbf{y}_h$ : the predicted concentrations (M×1)

$\mathbf{X}_t$ : the matrix of new objects, the test set (M×K)

**b** : the regression vector of Equation 1

As a way of testing the regression model, samples with known reference values ordered in a vector $\mathbf{y}_t$ (corresponding to the rows in $\mathbf{X}_t$) are used. This gives a prediction residual:

$$\mathbf{f}_t = \mathbf{y}_t - \mathbf{y}_h \qquad (3)$$

When the test set is made from mean spectra representing regions or entire images, the resulting predictions represent a mean region or global value. The values obtained in this way are effectively the numerical averaging equivalent to the physical averaging effect obtained from a spot probe or single beam type spectrometer. However the same regression models can also be applied to the individual spectra within a specified region or image, thereby providing prediction values at every pixel. This is one of the major benefits of hyperspectral image regression.

Equations 1 and 2 were first developed and applied to spectral data measured with a conventional spectrometer to enable prediction of quantitative values associated with individual samples. Regression of hyperspectral images provides an additional advantage in that **y**, $\mathbf{y}_h$ and $\mathbf{f}_t$ also represent images or parts of images and consequently can be interpreted visually by using their respective spatial indices, ij. In addition to providing spatial concentration profiles, examination of prediction and residual images from both calibration and test sets can yield important visual information regarding the proper choice of the number of latent variables.

### 2.3 Regression diagnostics

Diagnostic inspection of the calibration and prediction results is necessary to validate regression models and identify problems. As with conventional spectroscopy, a plot of the known vs. predicted values, **y** against **Xb,** should be examined. Ideally all plotted points should lie on a straight diagonal line. Nonlinearities, outliers, clustering of data points, heteroscedasticity, high leverage points, and errors in the design of **X** and **y** may be detected in such a plot [13]. Root mean square error of the calibration set (RMSEC), root mean square error of cross validation (RMSECV) or root mean square error of prediction of the test set (RMSEP) can be computed from residuals. PRESS plots (Predicted Residual Error Sum Squares) may assist in determining the proper number of components to use in the calibration model. These diagnostics are indicative of errors in the calibration model or the predictive ability of the

model applied to the entire test dataset.

When the calibration and test sets are based on small sets of spectra representing the mean spectra of regions or images, the resulting diagnostic values are expected to be similar to those obtained from equivalent spectra acquired from spot probe instruments. This approach allows for instrument comparisons, but provides no clear advantage for hyperspectral images. But hyperspectral images offer an additional diagnostic benefit. Since each image may contain a very large sample or test set population, accurate population statistics can also be estimated from predicted values.

Two descriptive parameters can be easily estimated to characterize the population of residuals obtained from the set of predictions representing a single image or region of interest: the population bias $r_n$ or average residual, and $s_{n,}$ the estimated standard deviation of the prediction residuals. If each region or image contains a true homogenous population, then these estimates represent the precision and accuracy of the model for predicting the constituent within the individual region. In this case the standard deviation estimate also describes the uniformity of the constituent within the region or image, a measurement which is clearly not obtainable from a spot probe technique.

Additional diagnostics can be formulated to compare prediction results within an image as well as between images. Residual means and standard deviations may be compared directly between regions or images, or pooled, resulting in global diagnostic values. A global RMSEP can be computed based on the average residual or bias from each image:

$$RMSEP_G = [\ \mathbf{r}^T\mathbf{r} / P\ ]^{½} \quad (4)$$

**r** : vector of bias values for each region
Similarly a pooled standard deviation can be computed from the individual standard deviations:

$$S_G = [\ \Sigma(s_i^2\ (L_i\text{-}1)) / (\Sigma\ (L_i - 1)\ ]^{½} \quad (5)$$

summations: i = 1 to P

When large regions of equal size are specified, Equation 5 may be further simplified:

$$S_G \approx [\ \Sigma(s_i^2\ (L))/(\Sigma(L)\ ]^{½} = [\ L\Sigma s_i^2\ /(\ L\ P)\ ]^{½} = [\ \mathbf{s}^T\mathbf{s}/P\ ]^{½} \quad (6)$$

**s** : vector of standard deviation values for each region

$S_G$ represents the precision or level of variance expected within one single image, while $RMSEP_G$ is a mixed measure of accuracy and precision representing the combined variance from all images. When regression models are based on a reduced space as in PLS, it may be useful to examine plots of $RMSEP_G$ vs. $S_G$ as a function of the number of latent variables included. For an underfit model $RMSEP_G$ may be quite large. As additional latent variables are added, the overall fit improves and $RMSEP_G$ is reduced. However any additional variance included in the model that does not correlate with Y, will contribute to an increase in the variance of the predicted values, $S_G$. A new metric which is based on both $RMSEP_G$ and $S_G$ can be computed:

$$D_G = [\ w_1\ RMSEP_G^2\ +\ w_2\ S_G^2\ ]^{½} \quad (7)$$

$w_1$, $w_2$: are balancing weights. In this paper $w_1 = w_2 = 1$.

When the two weights are both 1, this new metric is effectively the distance from the plot origin to the points plotted in the plots of $RMSEP_G$ vs. $S_G$. A plot of $D_G$ vs. the number of latent variables can be used to determine an optimal number of model components. This diagnostic incorporates variance contributions due to differences between samples imaged, sample variation within each individual region, and instrumentation changes. For model comparison purposes, the images should be well conditioned: instrument variation should be compensated for with the use of internal standards, and calibration set samples should be as uniform as possible throughout the region of interest.

## 3. EXPERIMENTAL/MATERIALS AND METHODS

### 3.1 Hyperspectral Images

Near infrared (NIR) hyperspectral images were acquired using a Spectral Dimensions MatrixNIR [14] camera, producing images with 256 x 320 pixels at 118 wavelength channels, 960 – 1662 nm, with 6 nm resolution. Images were digitized with 12 bit resolution, with an average of 10 scans recorded. A 64 ms integration time was used. The image field of view was 50 x 62 $mm^2$ with approximately 0.2 x 0.2 $mm^2$ pixels. Samples were either placed in circular NIR sample holders with glass covers, or placed on pieces of silicon carbide sandpaper. The sandpaper is inexpensive, disposable, and produces a very low reflectance background. Image acquisition time was 3 – 6 minutes so samples which were heat sensitive were cooled with a small box fan positioned just beyond the field of view.

The hyperspectral images were acquired as detector signal counts which needed to be transformed to reflectance before further processing. A series of five Spectralon [2] standard reference materials at 2, 25, 50, 75 and 99 percent reflectance were first imaged, allowing creation of independent reflectance transform functions at every pixel at every wavelength. Reflectance is commonly computed from single point transform functions based on only a 99 percent measurement and an instrument ‘dark’ or background measurement. The full range of Spectralon image hypercubes permitted the use of first or second order transform functions without the need to specifically measure background.

A calibration gray scale made from a second set of small Spectralon pieces was positioned to appear along the edge of every hyperspectral sample image. These standards permitted additional corrections or instrument standardization. A detailed description of these external and internal standards and their modeling algorithms has been given in a previous publication [1]. Figure 1a shows a typical image at one wavelength, including the internal standard gray scale. The imaged field of view was illuminated by 4 quartz halogen lamps. To test the robustness of the standard correction techniques, sets of Spectralon and sample images were acquired with the

lamps positioned for both maximally uniform and less ideal non-uniform illumination intensities.

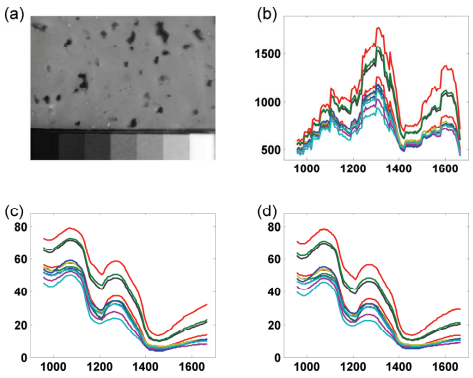


**Figure 1. Hyperspectral images viewed as image channels or spectra: (a) An image representing a single channel including an internal standard gray scale. Median spectra from ten cheese samples in raw instrument counts (b), transformed to percent reflectance using a single point transform (c), or second order transform (d). The horizontal axis for (b), (c), and (d) is wavelength in nm.**

### 3.2 Datasets

An artificial three component sample set was made from reagent grade citric acid (CAS 77-92-9), salicylic acid (CAS 69-72-7), and sugar (consumer grade Dan Sukker). All three pure chemicals were first screened with a 25 µm sieve to remove large particles, and then blended using a three component third order augmented simplex experimental design yielding 3 pure component, 6 binary (1/3, 2/3) and 4 ternary (1/3, 1/3, 1/3; 2/3, 1/6, 1/6; etc.) mixture samples. This collection of samples was designed to examine the detection limits and prediction abilities of different PLS calibration models and data pretreatments. The salicylic acid has a relatively unique spectrum with a strong peak at 1130 nm whereas the sugar and citric acid spectra are more collinear with overlapping peaks at 1438 and 1470 nm respectively. Samples of each mixture were placed in Foss NIRSystems [15] sample holders and imaged. The 3 pure and 6 binary mixture samples were imaged in duplicate, the 3 augmented ternary (2/3, 1/6, 1/6) mixtures in triplicate. For determining modeling statistics the center point ternary mixture (1/3, 1/3, 1/3) was imaged a total of 10 times. A single set of mixture samples without replacement was used. Replicate images were made in random sequence with random orientation of each circular sample holder. Two of the binary mixture replicates were removed from the dataset due to instrumentation failure, leaving a total of 35 hyperspectral images. A circular mask with radius of 80 pixels was used to identify the sample ROI, providing over 20,000 spectral samples from each hyperspectral image.

A second dataset was obtained using an assortment of thirteen commercial cheese products, listed in Table I. These cheeses were specifically selected to span as large a range as possible in concentrations of protein, fat, and carbohydrate. Initially the published values on the packaging labels were used as standard reference values. A parallel set of values for protein and fat content was determined using standard techniques AOAC 976.05 (Kjeldahl protein determination) and BS EN ISO 1735:2004 (fat extraction). Triplicate measurements of 5 cheeses yielded a standard deviation of 0.14 percent protein, and 0.41 percent fat. For imaging purposes, cheese slices 2-4 mm thick were positioned on top of the black silicon carbide sandpaper to minimize background contributions. Alternatively round cheese cores were cut and placed inside the Foss NIR sample holders and imaged through quartz glass windows.

Sets of images were acquired on five different days with varying numbers of replicates each day, providing a total of 80 images. One set included 12 replicate images of three cheeses acquired after purposely reducing the lamp power from the original 64 percent to 63, 62, 61, and 60 percent. These 12 images were excluded from global calibration and test sets. The image set from one of the other days was partitioned into two subsets because of an instrumentation disturbance: the liquid crystal tunable filter (LCTF) was reprogrammed giving a slight change in wavelength bandwidths. Both subsets were included in the global image sets.

Table I Protein, fat, and carbohydrate composition of cheeses from packaging labels and analysis[†].

| Cheese | Protein[*] | Fat[*] | Carbohydrate[*] | Protein[†] | Fat[†] |
|---|---|---|---|---|---|
| 1 | 26.0 | 31.0 | 1.5 | 24.9 | 32.6 |
| 2 | 30.0 | 17.0 | 1.5 | 29.8 | 16.8 |
| 3 | 27.0 | 28.0 | 1.5 | | |
| 4 | 32.0 | 10.0 | 1.5 | | |
| 5 | 24.0 | 33.0 | 1.5 | 27.1 | 35.3 |
| 6 | 32.0 | 5.0 | 2.5 | 33.2 | 5.7 |
| 7 | 31.0 | 10.0 | 1.0 | 32.8 | 11.7 |
| 8 | 25.0 | 26.0 | 1.5 | | |
| 9 | 13.5 | 18.5 | 7.6 | 14.4 | 17.9 |
| 10 | 10.5 | 26.3 | 10.5 | | 25.7 |
| 11 | 18.0 | 12.0 | 6.5 | 18.1 | 12.4 |
| 12 | 17.5 | 20.0 | 2.9 | 17.0 | 18.9 |
| 13 | 29.0 | 23.0 | 1.5 | 26.4 | 22.3 |

[*] Reference values from packaging labels; [†] Reference values from replicate chemical analysis.

It is important to note that while a single hyperspectral image may provide twenty to eighty thousand sample spectra, there were only 13 actual cheese samples measured. Additionally, while the protein content ranged from 10.5 to 32% and the fat content from 5 to 33% the expected carbohydrate concentration was mostly 1.5% with only a few special cheeses being higher. This made the distribution skew and it may have caused problems with cross validation. These are serious challenges to forming robust and accurate regression models.

**3.3 Data Pretreatments and Analysis**

Some of the sensors in the imaging system produced obvious hardware dead pixels with consistent signal counts of zero or 4096, the minimum or maximum A/D converter value. Other pixels had saturated values due to specular reflection or simply provided erroneous data. An iterative pixel removal process was used to eliminate these pixels from each hyperspectral image: a general spatial region of interest (ROI) was first identified based on sample geometry and further masked based on a simple threshold level. From this subset class of spectra, the standard deviation of all Euclidean distances to the median ROI spectrum was computed, and any pixels with a distance greater than 5 standard deviations were eliminated. A new median spectrum, sorted distances and standard deviation were computed and used for additional outlier detection. This cleaning process was repeated, until no more pixels were removed. The resulting clean class mask was used for all subsequent sample pixel selection steps. Each of the images measured contained only one sample mixture, hence all the pixels in the clean mask were expected to belong to a single class population.

For PLS calibration modeling purposes a single median spectrum was first computed from the set of all spectra contained within the cleaned ROI mask of each of the calibration set images. This reduced the effect of variance within an image due to physical diffuse reflectance processes such as light scattering and removed the high leverage effect due to any extreme outliers not removed by the mask cleaning process. Extreme outliers have a large effect on mean value computations, consequently median spectra which are more robust to this effect were computed from the ROI masks. It should be cautioned that it is possible that the median spectrum values at each wavelength channel may in fact have come from different pixel or spatial locations. In this example, due to the very large sample population sizes, typically 20 to 60 thousand, this was not likely to have caused errors.

Two approaches towards spectral test set selection were used. To test the effectiveness of using the instrument in a spot probe mode, single median spectra were computed for each of the test image ROIs. Alternatively, to examine the spatial distributions afforded by hyperspectral images, complete sets of ROI spectra were included as test sets. In this case univariate tail trimming methods were employed before computing modeling statistics. Predicted values from the complete set of mask selected spectra were sorted and the first and last 2.5 percent of all results were removed before further computations of RMSEP, mean, or standard deviation. This trimming of ROI results was also justified by the massive number of samples within a typical image ROI.

Unless noted otherwise, all PLS models were based on a leave-one-out sample cross validation method: for each mixture or cheese sample, all replicate images were eliminated from the calibration set and a PLS model computed. This avoided any aliasing effects that would otherwise be introduced by including the replicate images in a leave-one-out image cross validation. In addition to the reflectance transforms and internal standard corrections, spectra were also converted to first and second derivative spectra using Savitzky-Golay filters [16-18] or converted from reflectance to absorbance. No variable or wavelength selection was performed; the entire spectral range was used for all computations. All analysis was performed using MATLAB version 6.1.

## 4. RESULTS AND DISCUSSION

**4.1 Calibration standards**

Figure 1 shows NIR spectra measured for a selection of ten cheeses: b) the raw instrument spectra; c) reflectance spectra using the single point transform; and d) reflectance spectra using the second order transform. Converting the spectra to reflectance corrected for the wavelength sensitivity of the detector sensors and wavelength variation of light source intensities. When viewed side by side, differences between the single point and second order reflectance transformed spectra were difficult to see. Closer examination of enlarged spectral plots suggested that the second order transform model generally performed best, yielding spectra which were smoother with slightly improved peak shape and more pronounced amplitudes.

For each of the sample images collected, median spectra from ROIs of each of the internal standards were used to validate instrument performance and used to correct for short term drift and other instabilities. Figure 2a shows the variation seen for internal standard number two (72 percent reflectance) in the same series of cheese images as Figure 1. The results of applying different correction models can be seen in Figure 2b - d. Correction models were based on either the maximum reflectance internal standard only (single point) or the set of all internal standards (linear and second order). Individual correction models were computed for each image, and subsequently applied to all spectra within the same image. The second order correction model based on all internal standards performed best at converging the individual spectra towards a target spectrum, and hence converging the individual images (in essence a slave instrument) towards a single target image, or master instrument. This was evident from comparing the range in the minimum and maximum reflectance values for each transform. A pooled variance was computed from all wavelengths of the internal standard spectra displayed in figure 2. The computed standard deviations for the uncorrected, and single point, linear, and second order corrected spectra were 0.64, 0.32, 0.16, and 0.12 respectively (percent reflectance), again indicating optimal performance from the second order correction transform. In the example provided in Figure 2a, one of the initial reflectance spectra showed a significant deviation with a peak at 1250 nm. Because this abnormality was measured in all internal standards, the correction procedure effectively removed it from the corrected spectra of Figure 2d. Further implications of the use of internal standard corrections are discussed below.

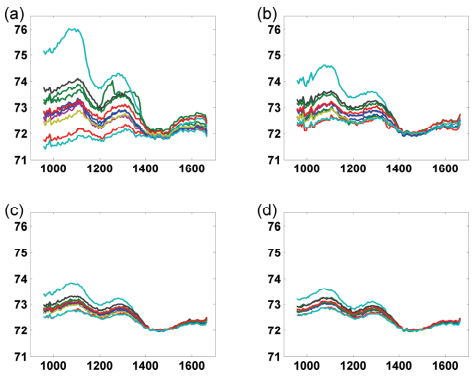


**Figure 2. Ten replicate spectra from the 72 percent reflectance standard (a) corrected by single point (b), linear (c), and second order (d), internal standard correction models. All plots are percent reflectance vs. wavelength in nm.**

### 4.2 Ternary mixture analysis

*4.2.1 Experimental design verification*

Principal component analysis (PCA) was performed on the mean centered median reflectance spectra obtained from the masked ROI of each of the replicate artificial mixture images. The $\mathbf{t}_1 - \mathbf{t}_2$ score plot is presented in Figure 3. The linear relationship of the mixing triangle was retained over the entire range of 0 – 100 percent concentrations of each constituent. The overall precision was very good, as indicated by the replicate scores of the 35 sample images. The ten replicates of the center point mixture clustered in a very small area. Replicates scores of other mixture samples were barely distinguishable from each other in the $\mathbf{t}_1 - \mathbf{t}_2$ space.

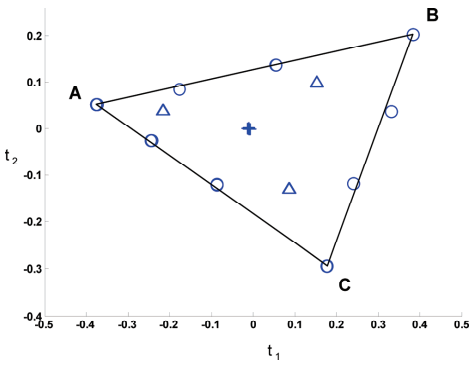


**Figure 3. PCA score plot of median spectra from 35 artificial mixture images. Duplicate constituent and binary mixture (circle); triplicate ternary mixture (triangle); and ten center point mixture (cross) replicates. The mixing triangle represents mixtures of sugar (A), citric acid (B), and salicylic acid (C).**

*4.2.2 Supervised exploratory analysis*

The immense quantity of data contained within a single hyperspectral image necessitates the use of chemometric data compression techniques to assist with image interpretation. This same volume of data also enables the extension of univariate tools to further data exploration. In the example dataset, a global PLS model was computed from the median spectra of all 35 mixture images, then applied to over 20,000 individual pixel or sample spectra measured in the masked region of one of the center point mixture images. Here K=118, N = 13, M=L≈20000. For each constituent of the mixture, the prediction results for the 20,000 spectra were summarized with a univariate histogram. These distribution plots offer insight into both the nature of the individual sample image, and the effectiveness of the regression model. The distribution shape, width, and any outlier locations provide information regarding the complexity of the sample, confirming sample purity or uniformity. The peak center location represents the average concentration and indicates model error.

Prediction histograms based on a four component PLS model are displayed in Figure 4 for the three mixture constituents. Although this was intended to be a well mixed uniform sample, the range in predicted concentrations was very large. For each constituent the peak center or mean concentration was very close to the expected 33 percent. However the range in the distribution of concentrations at individual pixels was very broad, with values ranging from 0 to 60 percent. The salicylic acid distribution was much narrower, probably due to the improved modeling of the more selective pure constituent spectrum. The peak shapes appeared very similar to a normal distribution, indicating that although there was significant variation from the expected concentration value, these deviations were randomly distributed.

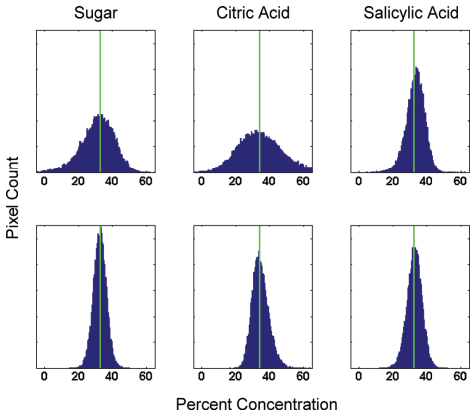


Figure 4. Histograms of 4 component model concentration predictions for a single center point mixture image. Model and predictions are based on absorbance spectra (top row) and first derivative spectra (bottom row). The expected concentration is indicated by the vertical line.

Table II Prediction errors and (number of components) for various diagnostic tests.

| Technique | Test Set | Sugar | | Citric Acid | | Salicylic Acid | |
|---|---|---|---|---|---|---|---|
| RMSEC | Median | 1.8 | (4) | 1.0 | (5) | 0.5 | (4) |
| RMSECV | Median | 4.0 | (4) | 4.1 | (5) | 2.6 | (2) |
| RMSEP - center point replicates | Median | 0.7 | (3) | 1.5 | (3) | 0.5 | (2) |
| $RMSEP_G$ - all images | ROI Set | 0.5 | (3) | 1.1 | (3) | 0.3 | (3) |
| $S_G$ | ROI Set | 2.6 | (3) | 3.4 | (3) | 3.0 | (3) |
| D-Metric | ROI Set | 2.7 | (3) | 3.6 | (3) | 3.0 | (3) |

The bottom row of Figure 4 contains histograms for the same sets of spectra, pre-processed by a first derivative transform. In this case, the use of a first derivative transform greatly decreased the variance in predicted concentrations. Similar broad normal distributions with a narrowing of first derivative based prediction histograms was observed for predictions from images of the pure constituents as well. This suggested that the first derivative transform corrected for a physical process artifact in all samples, most likely associated with light scattering or specular reflection.

Figure 5 demonstrates another modeling consequence observed in the prediction distributions, in this case sugar concentration based on first derivative spectra. As the number of components in the calibration model was increased from one to six, the distribution center shifted towards the expected value (decrease in bias) but the distribution also began to broaden (increase in standard deviation). This balance of distribution bias vs. distribution width is quantified in a new diagnostic descriptor for determining the number of latent variables, discussed further in the following section.

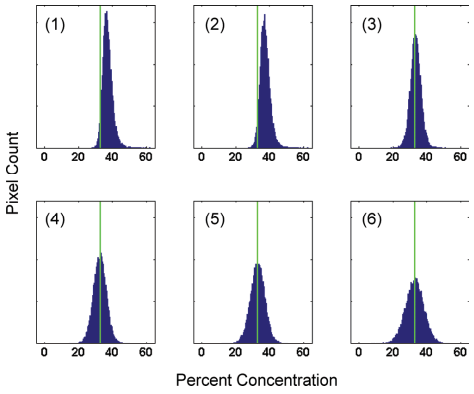


**Figure 5. Histograms of sugar concentration predictions from first derivative spectra of a single center point mixture image vs. 1 – 6 PLS components in the calibration model.**

The visual inspection of the population distributions and statistics is a major analysis tool that takes advantage of the additional information contained within a hyperspectral image. The broad range in predicted values and impact of the first derivative transform, and the change in width of distribution width as a function of the number of model components, would not be easily detectable from the single spectra that a single beam or spot probe spectrometer provide.

*4.2.3 Number of components*

A critical step in the creation of accurate regression models is choosing the correct pseudorank or number of latent variables to use. Many classical chemometric approaches have been suggested and were considered. These tests were based on either calibration model statistics or statistics based on predictions of test set spectra. These various tests are summarized in Table II, for first derivative reflectance spectra only. In all cases, median image spectra were used for creating calibration models. Test sets were based on either the median spectrum from each image or the collection of all spectra within each image ROI ($L_i \approx 20000$).

The first test was based on RMSEC, the error in predicting all samples, using a calibration model based on all samples. (K=118, N=35, M=35) This diagnostic suggested that 4, 5, and 4 components should be used in the calibration model for the three constituents. Since all test spectra were used for creating the model the prediction errors were relatively small.

The diagnostic from leave one out cross validation, RMSECV, (K=118, N=35-Q, M=Q, P=35) indicated much larger prediction errors than the RMSEC. The optimal number of components for sugar and citric acid was similar, but the results suggested a significantly fewer number of components were needed for prediction of salicylic acid concentration. Since the center point mixture was imaged ten times, RMSEP was computed for this class of mixture samples alone as a pure test set diagnostic. (K=118, N = 25, M=10) This diagnostic suggested that 3 components were needed for sugar and citric acid and only 2 components were needed to accurately predict the salicylic acid. The primary difference with this test is the dramatic reduction in prediction errors, especially for sugar and salicylic acid, to values less than 1 percent.

Because predictions from hyperspectral images offer large population statistics, diagnostic tests were included to take advantage of both inter and intra image statistics. From the 35 total images, a pooled standard deviation $S_G$ (Equation 6) and global $RMSEP_G$ (Equation 4) were determined from the individual standard deviation and bias values computed from the prediction populations for each individual image. (K=118, N=35-Q, M=L≈20000, P=35) These global values utilized the bias and standard deviations of the

distributions as in Figure 5, but were summarized for all images. The values resulting from 3 component calibration models are listed in Table II. The $RMSEP_G$ values were very similar to the RMSEP values for the predictions of the center point median spectra, ranging from 0.3 to 1.1 percent, and represented an overall prediction error resulting from the global effect of averaging all spectra predictions. However within an image the pooled standard deviation suggested that the average prediction error was close to 3 percent for all constituents. As indicated in the histograms of Figures 4 and 5, the PLS models do a good job of predicting the mean image concentration (< 1 percent), but there is significant variation (3 percent) in the predicted concentrations within the sample ROI of each image.

The variation of $RMSEP_G$ and $S_G$ is displayed in the scatterplot of Figure 6, as a function of the number of latent variables in the calibration model for the three ingredients. It is clear in this plot that as the number of components increases, $RMSEP_G$ decreases, indicating increased accuracy. However when too many components are included, the standard deviation increases, meaning less precision. The objective in this plot is to minimize both the $RMSEP_G$ and $S_G$ simultaneously. This can be done visually, or by computing the distance to the origin using Equation 7. Using this metric, the optimal number of model components suggested was 3 for all constituents.

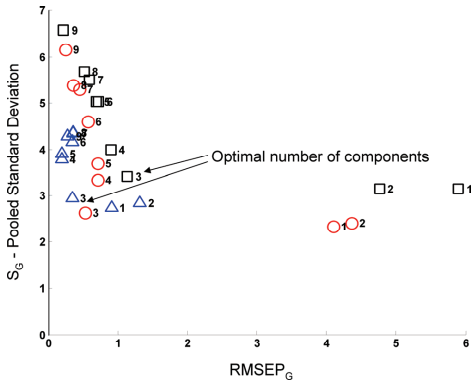


**Figure 6. Center point mixture concentration prediction errors from first derivative spectra: RMSEPG vs. SG, the pooled standard deviation of errors from 10 replicate images, sugar (circle), citric acid (square), salicylic acid (triangle). The number indicates the number of components in the calibration model.**

This combined metric includes variance contributions from cross validated calibration models, as well as large test set population statistics, taking full advantage of the complete hyperspectral image information. Press plots of this metric allow summary comparisons of trends as a function of the number of latent variables or differences between ingredients. The resulting distance value is representative of the prediction error expected in all images.

*4.2.4 Prediction Image Maps*

Hyperspectral images offer an analysis benefit in addition to the large population statistics obtained from the large number of sample spectra: spatial information. The sets of concentration prediction results can be viewed not only as histograms, but the individual prediction results can also be mapped back to the original spatial locations of the source spectra used for prediction. These spatial maps can then be viewed as an image, indicating the spatial distribution of ingredient concentrations. The spatial mappings of sugar, citric acid, and salicylic acid for one of the center point sample images are provided in Figure 7. The apparent high contrast texture in Figure 7 a and b, explained the broad range of values observed in the concentration histograms of Figure 4. The salicylic acid concentration had a much narrower histogram, which was confirmed by the smoother appearing concentration map of Figure 7 c.

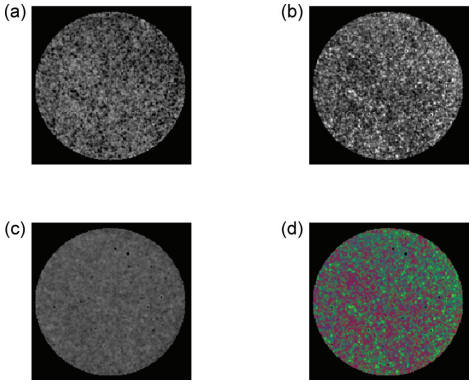


**Figure 7. Concentration prediction maps for a single center point image. Sugar (a), citric acid (b), salicylic acid (c), and a false color composite of all three constituents (d). Approximately 20,000 pixels or sample spectra in circular mask with 80 pixel radius.**

When viewed individually, the concentration maps for sugar, citric acid, and salicylic acid appear to exhibit a fairly uniformly distributed variation (uniform detailed 'texture') with a very subtle gross variation (slightly darker areas.) The three images can also be overlaid as red, green, and blue layers producing a false color image. Figure 7 d represents the resulting false color image after first scaling the predicted concentration information. The 0 to 100 percent possible range of concentrations was scaled to 0 to 1.0, for each ingredient, sugar (red), citric acid (blue), and salicylic acid (green). Again the fine structure variance or 'texture' appears fairly uniform, however now large areas (color blotches) of different relative concentrations of the three ingredients are visible. The uneven coloring of Figure 7 d was possibly due to constituent agglomeration within the 0.2 x 0.2 $mm^2$ pixel areas. Although the 25 micrometer sample particle sizes should have been much smaller than the individual pixel size, physical forces between particles may have prevented uniform mixing of the three constituents. The false color mapping technique is

Table III Cheese prediction errors for different data preprocessing techniques. Four component PLS-1 models are based on single median spectra from each image ROI.

| Technique | Protein* | Fat* | Carbohydrate* | Protein† | Fat† |
|---|---|---|---|---|---|
| RMSEC: | | | | | |
| Reflectance | 2.7 | 2.4 | 1.6 | 2.5 | 2.4 |
| Reflectance - First Derivative | 2.3 | 1.3 | 1.4 | 2.0 | 1.4 |
| Reflectance - Internal Std Corrected | 2.9 | 2.3 | 1.6 | 2.7 | 2.4 |
| Absorbance | 3.1 | 2.8 | 1.9 | 2.6 | 3.0 |
| RMSECV: | | | | | |
| Reflectance | 4.0 | 3.1 | 2.7 | 3.5 | 3.2 |
| Reflectance - First Derivative | 4.0 | 1.7 | 2.5 | 3.5 | 1.8 |
| Reflectance - Internal Std Corrected | 4.0 | 2.9 | 2.5 | 3.6 | 3.2 |
| Absorbance | 4.6 | 3.6 | 2.8 | 3.8 | 4.0 |

* Reference values from packaging labels; † Reference values from replicate chemical analysis.

extremely useful for providing additional sample uniformity information.

*4.2.5 Additional Transformations*

An extensive comparative analysis was performed on the images using the various data pretreatment techniques described: the simple, linear and second order reflectance transforms, and the internal standard corrections, all in combination with or without first derivative or absorbance transformations, as well as first and second degree spectra normalization. With the exception of the first derivative transform, these pretreatments had very little systematic effect on prediction results. In fact regression models based on the raw instrument count data performed nearly as well as the second order reflectance transform with second order internal standard correction. However as indicated in Figures 1 and 2, these transforms and corrections greatly enhance the interpretability of the spectra and are useful for instrument validation and performance diagnostics.

In this artificial dataset, sample presentation and instrument errors far outweigh the benefits available from data pretreatment and correction. Splitting the replicate images into different collections of calibration and test sets created substantial differences in the prediction errors and the determination of the optimal number of components. This was observed with test sets made up from both ROI median spectra and complete ROI spectra. Trends in RMSEC and RMSECV press plots changed dramatically for different calibration – test set splits with identical preprocessing conditions. The results presented here are based on the complete set of all images, and are valid for comparative purposes. More robust models would be expected from a larger set of replicate images acquired from multiple samples of the artificial mixtures. After examining the results of many different calibration-test set splits, it appeared that the D-metric seemed to be most robust to the effects of sample variation. In most cases the measured error was around 3 percent for all constituents. This reflects the variation of concentration within an image ROI.

**4.3. Cheese Analysis**

*4.3.1 Prediction Error Diagnostics*

The analysis of cheese represents a legitimate test of the potential of hyperspectral NIR image regression. Initial problems were encountered with sample preparation and measurement. During the lengthy five minute imaging process the heat from the quartz halogen illumination lamps caused some cheeses to melt. This caused samples to move during image acquisition due to the heating stress and in some cases, fat separated forming small beads on the cheese surface. Minimizing the analysis time through instrument optimization and the addition of a small box cooling fan permitted acquisition of acceptable sets of images. Sample preparation and stability are nonetheless critical elements of hyperspectral imaging of biological samples.

A comprehensive study of PLS prediction errors in determining protein, fat and carbohydrate concentrations was performed using various image data pretreatments and corrections. Both reflectance and absorbance transforms were compared to the use of spectra in raw instrument detector counts. The effect of instrument standardization based on the internal standards was also explored, correcting each image to either a grand master target image, or a target image for each day. Median ROI spectra were computed from each set of ROI spectra after the preprocessing. As with the artificial ternary mixture datasets, it was difficult to find a pretreatment modeling strategy that consistently improved the prediction error or readily identified the optimal number of calibration model components to use. Regardless of the preprocessing conditions, between 2 and 4 components created a reasonable model for protein and fat with relatively similar prediction results. Table III lists some typical RMSEC and RMSECV values for a four component model of the entire set of replicate images acquired on five different days. (K=118, N=68, M=68 for RMSEC, and K=118, N=68-Q, M=Q, P=68 for RMSECV) Four components was not always the optimal number of components, but was chosen as a consistent value for comparison of data pretreatment techniques.

Table IV Global prediction error statistics based on four component PLS-1 models applied to complete ROI spectra test sets.

| Technique | Protein* | Fat* | Carbohydrate* |
|---|---|---|---|
| $RMSEP_G$ | 1.8 | 0.7 | 1.3 |
| $S_G$ Pooled standard deviation | 2.3 | 1.4 | 1.4 |
| D-Metric | 3.0 | 1.6 | 1.9 |

* Reference values from packaging labels.

Table V RMSEP values before and after internal standard corrections for two types of instrumentation disturbances.

| Technique | Protein* | Fat* | Carbohydrate* | Protein† | Fat† |
|---|---|---|---|---|---|
| Power reduction: | | | | | |
| Uncorrected | 4.3 | 6.9 | 1.2 | 3.4 | 6.9 |
| Global standardization | 3.1 | 2.4 | 1.0 | 1.7 | 2.6 |
| LCTF Reload: | | | | | |
| Uncorrected | 2.4 | 2.1 | 1.7 | 2.2 | 2.1 |
| Global standardization | 2.9 | 2.4 | 1.7 | 2.8 | 2.6 |

* Reference values from packaging labels; † Reference values from replicate chemical analysis.

The regression models based on the ROI median spectra were also applied to individual image ROI spectra to compute predictions at each pixel location. Summary $RMSEP_G$, $S_G$, and D-Metric values for the images acquired on one day are listed in Table IV. (K=118, N=12-Q, M=L≈67000, P=12) The results listed are for reflectance spectra without any further preprocessing.

The example results listed represent the errors found when using all images from all days, Table III, or only the images from one specific day, Table IV. It was very difficult to make any general performance conclusions valid for any specific set of conditions. As observed in the artificial ternary mixture analysis presented in the previous section, the choice of calibration set and test set images and the application of different data pretreatments provoked major changes in the general trends in the different RMSEC and RMSECV press plots. Again one of the more robust error diagnostics was the D-metric based on the global $RMSEP_G$ and $S_G$. Figure 8 shows the D-metric press plot for the same one day test set images used for Table IV values, and suggested that in this particular case only two components were sufficient for each ingredient.

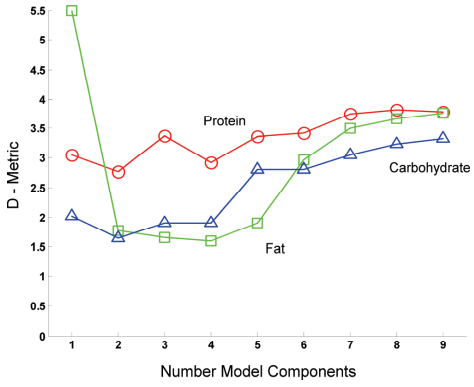


**Figure 8. D-metric combining RMSEPG and SG prediction errors in cheese. Protein (circle), fat (square), carbohydrate (triangle) vs. number of components in calibration model.**

The $RMSEP_G$ is a general indicator for how well the cheese constituents can be predicted. Typical errors of 1 – 2 percent for protein and fat, and 2 to 3 percent for carbohydrate were determined, which was greater than the chemical analysis error measured ( 0.14 percent protein and 0.41 percent fat) but similar to results obtained with other NIR spectrometers [19]. Unfortunately the reported value of carbohydrate in most of the cheeses sampled was 1.5 percent, which is below the indicated detection limit. The larger $S_G$ values indicated additional variation of constituent concentrations within an image or cheese sample. Models based on constituent reference values obtained from packaging labels gave comparable results to those obtained from models based on the reference values obtained from chemical analysis for protein and fat.

*4.3.2 Internal Standard Corrections*

For most of the images, the internal standard correction preprocessing step made insignificant changes to the prediction results. However two special situations were examined. A test set of 12 replicate images was made of three cheeses after purposely reducing the power supply voltage to the sample illumination lamps. A second test set of 9 replicate images was obtained after a reset in the LCTF wavelength bandwidth settings due to a thermal drift in the filter. Principle component analysis of median ROI spectra from the internal standards within the entire image set clearly showed these two instrumentation disruptions as isolated clusters.

The entire set of hyperspectral images was standardized by computing correction models for each individual image based on its internal standards and a common set of

standard target spectra. Corrected median ROI spectra were then computed and used as calibration and test set spectra. RMSEP results are presented in Table V for the lamp power reduction (K=118, N=59, M=12) and the LCTF reset (K=118, N=59, M=9) based on a four component PLS-I model consistent with previous examples.

The internal standard correction clearly improved the predictions from the reduced power images. The change in lamp power resulted in a change in reflectance levels throughout all spectra, which was detected and compensated for by use of the internal standards. The LCTF reset made subtle changes to the spectral bandwidths which were detected by the PCA of the internal standard spectra. However this variance was not well correlated to the constituent variances and little change in prediction error was observed. In either case the inclusion of the internal standards enabled detection and possible correction of the instrumentation disturbances.

*4.3.3 Prediction Image Maps*

In addition to the average ingredient values, the concentration predictions from individual hyperspectral image spectra can be mapped to provide spatial information. Figure 9 provides the false color maps for the 2 component prediction of protein, fat, and carbohydrate in one set of images for 12 individual cheeses. (K=118, N=12-p, M=p, P=12) For comparison purposes, the false color of the combination of expected values of the three ingredients was depicted in a smaller rectangle below each prediction image. Color intensities were scaled such that 0-40, 0-40, and 0-10 percent ranges of protein, fat, and carbohydrate, were mapped to the full scale values of red, green, and blue. A wide range of target colors was displayed, representing the broad spectrum of compositions depicted in the cheese dataset. The color matching between expected and predicted values was quite good. The false coloring also aided spatial analysis. Slight inhomogeneous areas were easily noticed due to changes in coloration of the images. Concentration prediction maps resulting from purposely choosing overfitting calibration models may also be viewed. The additional variance modeled may provide spatial sample and instrument performance information. In addition to the overall or average concentration values which are available with a probe type NIR instrument, the hyperspectral imaging provides detailed spatial resolution, which may be equally important in the manufacturing of a product such as cheese.

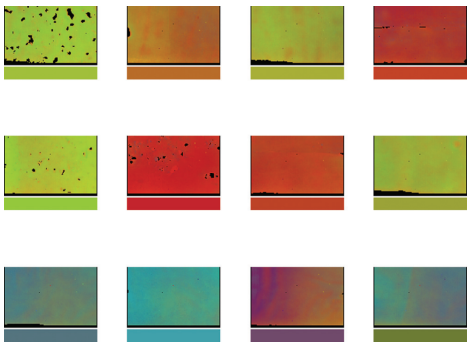

**Figure 9. False color concentration prediction maps for 2 component first derivative spectra models of 12 different cheeses. A smaller rectangle below each prediction map indicates the target 'color' expected. Rectangular cheese masks are 209 x 320 pixels.**

The images represented in Figure 9 were acquired with the illumination lamps purposely defocused. Large areas of the image received only 60 – 70 percent of the full illumination intensity. This intentional unevenness was effectively flattened or removed using the second order external standard reflectance transform, permitting continuation with regression analysis and prediction of ingredients.

## 5. CONCLUSIONS

Regression of hyperspectral image data provides a powerful tool for increasing the understanding of sample constituent concentrations and their spatial variation. One striking property of the examples in this paper is that hyperspectral images often offer large spectral populations but only small sets of reference values. If known spatial regions of interest within an image are mapped to a single reference value, the collection of spectra in that region of interest can be represented by a single median spectrum. Sets of these median spectra can then be used to build calibration models which in turn can be applied to large populations of test set spectra to obtain large populations of predicted values. Histograms and statistic estimates of these populations provide information regarding the uniformity of constituent concentration. Averaged quantitative results can be computed which are comparable to those obtained with spot probe spectroscopic techniques. Mapping the concentration values adds spatial information, allowing even greater insight into sample uniformity. This may be further accentuated by use of false color mapping visualization techniques.

Reliable prediction results can only be produced from accurate calibration models based on reliable data. It is important to ensure that only spectra truly representative of the known constituents are included in the calibration set. This is achieved by first constructing spatial region of interest masks, followed by the use of thresholding and within class distances to effectively remove all significant outlier pixels or spectra. It is also important to select the correct number of latent variables or components in the calibration model; too few or too many produce poor predictions. The availability of estimates of large population statistics of predicted values offers a new robust approach: a global RMSEP value provides a general measure of overall bias, while a pooled standard deviation describes the variance in prediction. A new distance metric has been proposed which balances contributions from these two measures.

In the two example datasets examined, the use of a first derivative transform did little to change the average predicted concentration value of constituents however it provided a major reduction in the range in predicted values. This was clearly observed as a peak narrowing in the histograms of predicted values and as a noticeable smoothing in the concentration image maps. Transformations of spectra from instrument detector counts to reflectance or absorbance had minimal impact on the

constituent predictions but are useful for spectral interpretation. Similarly image standardization by use of internal standards included within the image field of view had little effect on constituent prediction but is still recommended to allow instrument and experiment diagnostics. This is especially important when a standardized instrument image is needed to make use of cataloged library spectra. Considerable variation of prediction results was observed after selection of different combinations of calibration and test set spectra. Examination of the internal standard spectra of questionable images confirmed that the instrument was performing properly and not the source of this variance. Other sources of variance need to be investigated further. Physical effects such as light scattering can be mathematically corrected for. Sample preparation and presentation techniques of materials for quantitative and qualitative hyperspectral image analysis should be carefully examined to further reduce sample variance as much as possible.

**Acknowledgements**
The authors thank NIRCE, an EU Unizon-Kwarken project, Kempestiftelserna grant SMK-2062, and SLUP for financial support of this project. We also acknowledge Jan Sjögren, TestElek Svenska AB, for providing the Spectralon internal standard gray scale strip material, and Carina Jonsson for performing additional cheese fat and protein analysis.

**REFERENCES**

1. Burger J, Geladi P. Hyperspectral NIR image regression part I: Calibration and correction. *J. Chemometr*. 2005; **19**: 355-363.
2. Spectralon Diffuse Reflectance Targets, Labsphere, Inc., PO Box 70, 231 Shaker Street, North Sutton, NH 03260, USA.
3. Beebe K, Pell R, Seasholtz M. Chemometrics a practical guide. Wiley: New York, USA, 1998, 26-47.
4. Naes T, Isaksson T, Fearn T, Davies T. Multivariate calibration and classification. NIR Publications: Chichester, UK, 2002, 107-114.
5. Naes T, Isaksson T, Fearn T, Davies T. Multivariate calibration and classification. NIR Publications: Chichester, UK, 2002, 24-25.
6. Höskuldsson A.The H-principle in modelling with applications to chemometrics. *Chemometrics Intell. Lab. Syst.* 1992; **14** : 139-153.
7. Kalivas J, Green R. Graphical diagnostics for regression model determinations with consideration of the bias/variance trade-off. *Chemometrics Intell. Lab. Syst.* 2002; **60**: 173-188.
8. Grahn H, Szeverenyi N, Roggenbuck M, Geladi P. Tissue discrimination in magnetic resonance imaging: a predictive multivariate approach. *Chemometrics Intell. Lab. Syst* 1989; **7**: 87-93.
9. Geladi P, Swerts J, Lindgren F. Multiwavelength microscopic image-analysis of a piece of painted chinaware - classification and regression. *Chemometrics Intell. Lab. Syst* 1994; **24**: 145-167.
10. Winson M, Goodacre R, Timmins E, Jones A, Alsberg B, Woodward A, Rowland J, Kell D. Diffuse reflectance absorbance spectroscopy taking in chemometrics (DRASTIC). A hyperspectral FT-IR based approach to rapid screening for metabolite overproduction. *Anal. Chim. Acta.* 1997; **348**: 273-282.
11. Yu H, Macgregor J, Haarsma G, Bourg W. Digital imaging for online monitoring and control of industrial snack food processes. *Indust. Engr. Chem. Res.* 2003; **42**: 3036-3044.
12. Yu H, Macgregor J. Monitoring flames in an industrial boiler using multivariate image analysis. *Aiche Journal* 2004; **50**: 1474-1483.
13. Draper N, Smith H. Applied regression analysis. (2nd edition) Wiley: NewYork,USA, 1981, 141-149.
14. Spectral Dimensions, Inc., 3416 Olandwood Court, Suite 210, Olney, MD 20832, USA.
15. Foss NIRSystems, Slangerupgade 69, Postbox 260, DK-3400 Hillerød, Denmark.
16. Savitzky A, Golay M. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 1964; **36**: 1627-1639.
17. Steiner J, Termonia Y, Deltour J. Comments on smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 1972; **44**: 1906-1909.
18. Madden H. Comments on smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 1978; **50**: 1383-1386.
19. Burger J, Geladi P. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *the Analyst* (accepted).

# III

# Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples

**James Burger* and Paul Geladi**

***Received 13th April 2006, Accepted 27th June 2006***
***First published as an Advance Article on the web***
**DOI: 10.1039/b605386f**

A hyperspectral image in the near infrared contains thousands of position-referenced spectra. After imaging reference materials of known composition it is possible to build Partial Least Squares (PLS) regression models for predicting unknown compositions from new images or spectra. In this paper a comparison is made between spectra from a hyperspectral image and spectra from two spectrometers: a scanning grating instrument with rotating sample holders and an FT-NIR instrument utilizing a fiber-optic probe. The raw spectra and the quality of the PLS calibration models and predictions are compared. Two sample datasets consist of a set of 13 designed artificial mixtures of pure constituents and a selection of 13 sampled cheeses. The prediction error from the hyperspectral image spectra is between that of the two spectrometers. For a typical food sample, the average bias [and replicate standard deviation] was −0.6% [0.5%] for protein and −0.2% [1.3%] for fat. Comparable values for the best spectrometer were −0.2% bias for protein and −0.5% for fat. Some of the advantages of working with hyperspectral images are highlighted: the simultaneous exploration of representations of both spectral and spatial data, and the analysis of concentration profiles and concentration maps all contribute to better characterization of organic and biological materials.

## Introduction

The calibration and prediction of composition and other properties in organic and biological materials by near infrared spectroscopy are well established techniques.[1–6] A near infrared spectrum of a biological sample can be used for calculating concentrations of *e.g.* fat, protein, water, carbohydrates *etc*. Recent advances in technology enable the acquisition of near infrared (NIR) hyperspectral images. These images consist of a near infrared spectrum for each pixel allowing, in theory, the prediction of analyte concentrations at each pixel, leading to the creation of concentration images or maps.

Hyperspectral images, also known as hypercubes, are created by overlaying hundreds of single channel black and white or grayscale images, each representing a single band of spectral wavelengths. A typical commercial instrument[7] presented in this paper produces hypercubes with dimensions 256 × 320 × 128. This can be interpreted as 128 single channel images each with 256 × 320 pixels. Alternatively, this same hypercube can be viewed as over 81 000 spectra, each with 128 wavelength points. This huge amount of data poses data mining challenges, but also creates new opportunities.

As with conventional spectroscopy, chemometrics can be applied to extract relevant information relating to the spectral content, allowing sample classifications or quantitative determinations. Multivariate exploration and classification of multivariate images have been described earlier.[8–10] When additional quantitative information is available for calibration hypercubes, partial least squares (PLS) and other regression models can be created for predicting future test set hypercubes.[11–14]

An important question to ask is how this developing technology compares with other more conventional NIR spectroscopic techniques. What are the advantages or disadvantages of using a hyperspectral imaging instrument? This paper explores these issues by making a direct comparison between three commercially available spectroscopic techniques: a classical single beam scanning grating NIR spectrometer (SG), a fiber optic probe FT-IR spectrometer (FO), and an NIR hyperspectral imaging instrument (HI). The results are based on the analysis of two sets of samples: a set of artificial test mixtures, made according to a simplex lattice augmented mixing triangle, containing salicylic acid, citric acid, and sugar in known mass fractions; a set of 13 commercially produced cheeses containing varying amounts of protein, fat, and carbohydrate.

The two spectrometers measure light intensities over large volumes of sample resulting in a single spatially averaged spectrum for each sample measurement. The hyperspectral images represent detailed local information measured from light intensities obtained from individual small volumes, generating a large population of sample spectra. How should a calibration model that has useful prediction properties be developed from such a large set of spectra? Calibration models between near infrared spectra are often made by *e.g.* partial least squares regression, sometimes after spectral

*Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, SLU Röbäcksdalen, PO Box 4097, SE 90403 Umeå, Sweden. E-mail: jim.burger@btk.slu.se; Fax: +46 90 786 8799; Tel: +46 90 786 8798*

preprocessing. Performance diagnostics based on the regression model as well as predictions of independent data test sets can be used to judge the analytical usefulness of the method. Hyperspectral image calibration models for prediction and their diagnostics are presented and discussed. Some advantages of using images are pointed out and illustrated.

## Experimental

### Datasets

An artificial three constituent sample set (TriMix) was made from reagent grade citric acid (CAS 77-92-9), salicylic acid (CAS 69-72-7), and sugar (consumer grade Dan Sukker). A three component third order augmented simplex experimental design yielding 3 pure constituent, 6 binary and 4 ternary mixture samples was used. All three pure chemicals were first screened with a 25 μm sieve to remove large particles, and then blended according to this design scheme. Table 1 lists all samples and the number of replicates measured. The designed mixture ratios were binary (1/3, 2/3), ternary (1/6, 1/6, 2/3), and center (1/3, 1/3, 1/3). Replicate HI and SG measurements of each mixture were made from a single subsample, while the FO replicates were acquired by repeated measurement of the bulk sample containers.

A second dataset was selected from an assortment of commercial cheese products. These cheeses were specifically chosen to span as large a range as possible in concentrations of protein, fat, and carbohydrate. Initially the published values on the packaging labels were used as standard reference values. A parallel set of values for protein and fat content was determined using standard techniques AOAC 976.05 (Kjeldahl method) and BS EN ISO 1735:2004 (fat extraction). The percent protein content ranged between 10.5 and 32, fat between 5 and 33, and carbohydrate 1.5 and 10.5. The protein and fat concentration values were fairly evenly distributed however the carbohydrate values were skewed with 10 of the 13 being less than 3 percent. The number of replicate measurements for all cheese datasets is listed in Table 2. All replicate measurements were made on independent subsamples. Measurements were made on five separate days from either 12 or 10 of the different cheeses. Set V was imaged and then immediately measured with the SG instrument.

### Hyperspectral images (HI)

Near infrared (NIR) hyperspectral images were acquired using a Spectral Dimensions MatrixNIR camera.[7] The spectral range recorded was 960–1662 nm, with 6 nm resolution, producing hyperspectral images with 256 × 320 pixels at 118 wavelength channels. Each channel was scanned 10 times and digitized with 12 bit resolution, using a 64 ms integration time. The image field of view was 50 × 62 mm with approximately 0.2 × 0.2 $mm^2$ pixels. Total image acquisition time was 3–6 minutes so heat sensitive samples were cooled with a small box fan positioned just beyond the field of view. This lengthy acquisition time is due to the settling time required by the liquid crystal tunable filter (LCTF) used in the camera; changes in the number of replicate scans had little effect on total image acquisition time. All images were acquired in raw signal format representing A/D sensor counts and stored directly to disk. Image files were subsequently transformed to reflectance using models based on a series of calibration images.[15] The time required for acquisition of calibration images was 30 minutes, with an additional 7 minutes needed for computation of reflectance transform models using a 3.2 GHz P4 processor. The subsequent conversion of individual hyperspectral images from A/D sensor counts to reflectance required only 20 s of processing time. Samples for imaging were either placed in circular NIR sample holders with glass covers, or placed on pieces of silicon carbide sandpaper. The sandpaper is inexpensive, disposable, and produces a very low reflectance background.

**Table 1** Sample datasets for each instrument

| Instrument | Sample | Number of samples | Number of replicates | Number of spectra |
|---|---|---|---|---|
| HI | Constituent | 3 | 2 | 6 × 20000 |
| | Binary mix | 6 | 2 | 12 × 20000 |
| | Ternary mix | 3 | 3 | 9 × 20000 |
| | Center mix | 1 | 10 | 10 × 20000 |
| SG | All samples | 13 | 3 | 39 |
| FO | All samples | 13 | 5 | 65 |

**Table 2** Sample datasets for each instrument

| Instrument | Sample set | Number of samples | Number of replicates | Number of spectra |
|---|---|---|---|---|
| HI | I | 12 | 1 | 12 × 20000 |
| | II | 12 | 1 | 12 × 20000 |
| | III | 12 | 1 | 12 × 20000 |
| | IV | 10 | 1 | 10 × 20000 |
| | V | 10 | 1 | 10 × 20000 |
| SG | V | 10 | 1 | 10 |
| FO | VI | 12 | 3 | 36 |
| | VII | 10 | 5 | 50 |
| | VIII | 10 | 8 | 80 |

### Scanning grating spectrometer (SG)

The Foss NIRSytems 5000 is based on a scanning grating monochromator with a PbS detector.[16] The wavelength range selected was 1100–2498 nm with 2 nm resolution. Each spectrum was averaged from 32 scans. Samples were contained in glass covered sample holders which spun during signal acquisition. These same sample holders were also used for hyperspectral imaging. The results were given immediately in absorbance by the instrument. The acquisition time per sample including the scanning of a standard reference material and conversion to absorbance was typically <60 s per sample.

### Fiber optic probe spectrometer (FO)

Spectra were also acquired from a Bruker Matrix F FT-NIR instrument, equipped with a fiber-optic sampling probe.[17] This instrument uses an InGaAs detector with a working range of 4000–12 000 $cm^{-1}$. Spectra were collected at 4 $cm^{-1}$ resolution, but for instrument comparison purposes were displayed in wavelength units, 833–2500 nm. Each spectrum was an average of 32 scans. The total acquisition time was typically <30 s for each sample spectrum. The probe was inserted

directly into containers of the artificial sample set mixtures, or positioned directly perpendicular on the surface of the cheeses. These spectral results were also given immediately in absorbance.

### Data selection and pre-treatment

Some of the sensors in the hyperspectral imaging system produce obvious hardware dead pixels with consistent signal counts of zero or 4095, the minimum or maximum measurable detector sensor value. Other pixels may have saturated values due to specular reflection or simply provide erroneous data. Euclidean distances to a mean spectrum were computed and then thresholded to eliminate these outlier pixels from each hyperspectral image. A region of interest (ROI) mask was then created to specifically select these *cleaned* spectra (pixels) to be included in PLS calibration and testing. For PLS calibration modeling purposes a single median spectrum was first computed from the set of all spectra contained within the cleaned ROI mask of each of the calibration set images. This reduced the effect of variance within an image due to physical diffuse reflectance processes such as light scattering and removed the high leverage effect due to any extreme outliers not removed by the mask cleaning process. Extreme outliers have a large effect on *mean* value computations, consequently *median* spectra which are more robust to this effect were computed from the ROI masks.

For initial comparisons of the instruments, test set spectra consisted of single median spectra computed from each of the hyperspectral test image ROIs. Alternatively, to examine the advantages of the spatial distributions and large sample populations afforded by hyperspectral images, complete sets of ROI spectra were included as test sets. In this case univariate tail trimming methods were employed before computing modeling statistics. Predicted values were sorted and the first and last 2.5 percent of all results were removed. This trimming of ROI results is justified by the massive number of samples within a typical image ROI, typically 20 to 60 thousand.

Except where noted, all PLS models were based on a leave-one-out sample cross validation method: for each sample, all replicate spectra were eliminated from the calibration set and a PLS model was computed. This avoids any aliasing effects that would otherwise be introduced by including the replicate spectra in a leave-one-out spectrum cross validation. Spectra were also converted to first derivative spectra using Savitzky–Golay filters and in the case of the hyperspectral image they were first converted from reflectance to absorbance. For the first derivative computation, a second order polynomial was used with a window size of 5 (HI) or 17 (SG and FO) points. Raw spectra were acquired for different wavelength ranges on each instrument; however a trimmed common wavelength region between 1100–1662 nm was also used. Within this range, the different instruments generated spectra containing between 94 and 797 discrete data points. The common wavelength instrument comparison was performed after first integrating and averaging each spectrum into an equal number of 94 evenly spaced bins with 6 nm bandwidths.

Hyperspectral image exploratory software, JIMIA, was used to enable interactive exploration of both spatial or feature space and spectral or PCA score space. This software was written by one of the authors in Java version 1.4 with supporting libraries from ImageJ,[18] JAMA,[19] and JFreeChart.[20] All other data analysis was performed using MATLAB version 6.1.

## Results and discussion

### Artificial mixture

The absorbance spectra of three pure constituents used in the artificial mixture test set are shown in Fig. 1(a), measured on the hyperspectral imaging camera. Fig. 1(b) displays the spectra of sugar measured on the three different instruments: hyperspectral imaging camera (HI), scanning grating spectrometer (SG), and fiber optic FT-IR spectrometer (FO). A wavelength range of 1100–1662 nm which is common to all three spectrometers has been used for comparative display purposes. The individual spectra displayed from the three instruments each represent an *average* spectrum: with the HI data, a median spectrum from a spatial region of interest (ROI) mask was computed; the SG spectrum is measured from a spinning sample, physically averaging the sample; the FO spectrum results from averaging the contributions from each

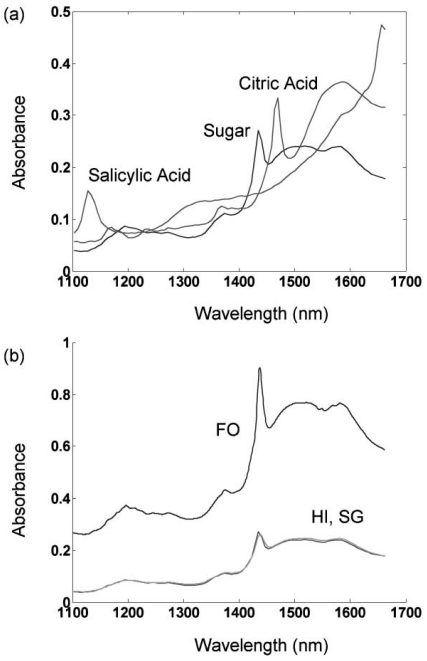


**Fig. 1** (a) Averaged hyperspectral image spectra of pure constituents salicylic acid, sugar and citric acid. (b) Sugar measured with hyperspectral image (HI), scanning grating (SG), and fiber optic probe (FO) instruments. The wavelength region displayed is common to all three instruments.

fiber in the probe bundle, also a physical averaging technique. The spectra displayed are the average values from replicate spectra measured from multiple samples on each instrument. This compound averaging effect minimizes any variance effects due to sample inhomogeneity or spatial sensitivity differences within each instrument.

Two important observations can be made from Fig. 1. First, the three constituents have unique spectral features: salicylic acid has a single peak centered at 1130 nm, whereas citric acid and sugar are more collinear, with overlapping peaks centered at 1470 and 1438 nm respectively. Secondly, the HI and SG spectra are nearly identical, however a significant spectral baseline offset was consistently observed with all measured spectra obtained with the FO spectra.

It is important to know how well the measured spectra represent the sample mixing triangle. Do the mixture spectra accurately reflect the constituent mixtures, or are there perhaps additional components contributing to the baseline differences? One technique of determining this is to use Principal Component Analysis (PCA) to project the spectral data into a reduced dimensional space. PCA was applied to the common binned–common wavelength range spectra from replicate measurements of each sample, for each instrument. Fig. 2(a) shows the resulting $t_2$–$t_3$ score space plots. The $t_1$ component effectively removes the average spectrum common to all spectra. The $t_2$ and $t_3$ scores represent the variance due to differences between the individual constituent spectra as seen in Fig. 1 as well as any additional sample and instrument variation. In all three instrument score plots the 13 points or distinct clusters of points of the designed experiment mixing triangle can be clearly seen. This is especially important with the FO score plot, indicating that the spectral offset observed in Fig. 1 occurs in all mixture spectra as well, and the offset amplitude is linearly proportional to constituent concentration. As with the raw spectra, the mixing triangle plots of the HI and SQ instruments agree quite well.

Fig. 2(a) also indicates the difference in relative precision between instruments. Scores from replicate spectra from all mixture samples are plotted. The SG instrument has the best precision and the scores of the triplicate spectra measured of each mixture sample can not be distinguished. The HI score plot shows only slight replicate variance; the ten replicated spectra of the center point mixture sample are very tightly clustered. In contrast however, the FO score plot shows a significant loss in precision; there is a significant variation within the cluster of each set of replicate mixture spectra. This is due to the fiber optic measurement carried out over a smaller sampling volume, and imprecision in inserting the probe into the sample matrix in a reproducible manner.

In an attempt to standardize the instruments, the absorbance spectra were converted to first derivative spectra before binning, and normalized to unit length. The resulting $t_1$–$t_2$ PCA score plot is displayed in Fig. 2(b). The three mixing triangles of the three instruments are now very similar, indicated by the score locations of the mixture spectra from all three instruments. This combination of transforms permits standardization of the three instruments, however it also destroys the linear relationships between absorbance and constituent concentration. This standardization technique was not used in the subsequent work discussed below.

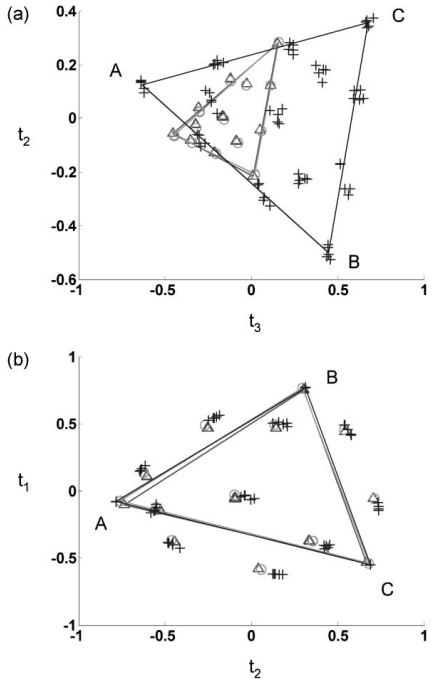


**Fig. 2** PCA scoreplots of replicate spectra, HI (triangle), SG (circle) and FO (cross), showing the designed experiment mixing triangle of sugar (A), citric acid (B), and salicylic acid (C). Data preprocessing treatments were absorbance (a) and first derivative + 2-norm (b).

PLS-1 models were computed using all available spectra except the center point sample mixture, which was used for testing model predictions. The sample replicate numbers are indicated in Table 1. While the optimal number of latent variables might vary with instrument or constituent, for comparison purposes, a four PLS component model was used for all predictions. Table 3 summarizes the results for all three constituents. It should be cautioned that these statistics are

**Table 3** Center point mixture predictions. Average bias and standard deviation of replicate measurements

| | Sugar | | Citric acid | | Salicylic acid | |
|---|---|---|---|---|---|---|
| | Bias | Std | Bias | Std | Bias | Std |
| Absorbance | | | | | | |
| HI | 0.6 | 1.2 | −2.8 | 1.1 | 1.7 | 0.2 |
| SG | 0.4 | | −0.7 | | 0.5 | |
| FO | 0.5 | 1.4 | −2.1 | 0.7 | 1.4 | 1.6 |
| First derivative | | | | | | |
| HI | 2.0 | 1.3 | −3.0 | 1.4 | 0.8 | 0.3 |
| SG | 0.1 | | −0.9 | | 1.1 | |
| FO | −0.5 | 1.7 | −1.9 | 0.6 | 2.1 | 1.7 |

based on very small numbers of samples, (10, 3, and 5 for the HI, SG, and FO instruments) however this variation represents typical uncertainties observed. The SG instrument provides the best precision, while the FO instrument the least. The bias generally ranged between $\pm 3$ percent, without observing any clear trends due to instrument differences. Results for modeling and predictions of first derivative spectra are also listed. No clear improvement in the predicted values was gained using this transformation.

These results are based on the modeling and prediction of the full spectral range collected from each instrument, which varied significantly from instrument to instrument. Computations were repeated, but restricted to the spectral range 1100–1662 nm, which was common to all three instruments. Additional predictions were made after first applying a first derivative transform to the absorbance spectra. Fig. 3(a) represents the results of all center point mixture sample predictions. No clear trend in differences was observed between using the full spectral range or the narrower common wavelength range, or between choice of absorbance or first derivative spectra. However the increased replicate sample variance observed in the FO spectra score plots of Fig. 2 is manifested in the large error bars of the FO predictions displayed in Fig. 3. The SG instrument provided the best precision, although these conclusions were based on only triplicate measurements made on the same sample of each mixture.

Because of different digital sampling techniques in the different instruments, the continuous spectral waveforms within the 1100–1662 nm range are digitized as 94, 282, and 797 discrete points for the HI, SG, and FO instruments respectively. For a more consistent comparison, the spectra from all three instruments were integrated and binned creating 94 point spectra with 6 nm bins. Additionally, the PLS loadings were examined to select 20 variables with significant loadings as a preliminary test of predictions based on wavelength variable selection. In all cases no clear trend was observed for improving the model–prediction process. The predominant source of observed variance came from the repeated measurements of the mixture samples. The accuracy in prediction was consistent between all three instruments, with most prediction errors less than $\pm$ 2 percent. The precision of replicate measurements was substantially different between instruments: the SG instrument offered the best repeatability, while the FO instrument provided the least.

### Cheese analysis

The artificial sample mixtures previously described are useful for testing the relative performance of the three spectroscopic instrument techniques with carefully designed laboratory test samples. However, most applied uses of spectroscopic measurements target real samples of organic and biological materials where the composition and structure are determined by nature or by food processing. An assortment of 13 cheeses was used to compare applied results on a complex food product. Cross validated leave-one-out models were used to first predict the full range of protein, fat and carbohydrate concentrations provided by the cheese manufacturers. Leave-one-out modeling was based on cheese type, that is any

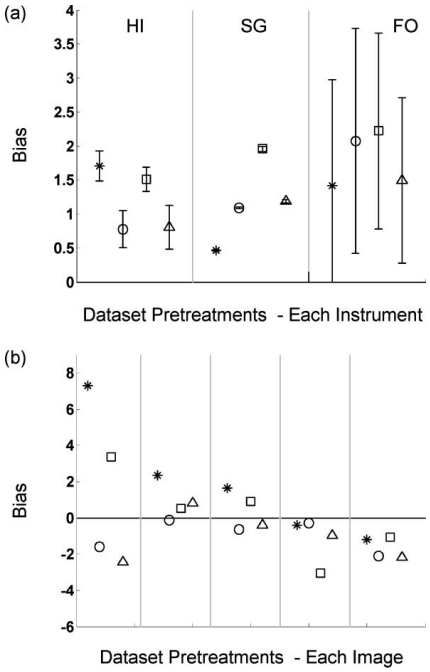


**Fig. 3** (a) Average bias with replicate standard deviation for a 4 component model of salicylic acid predicted from hyperspectral image (HI), scanning grating (SG), and fiber optic probe (FO) spectra. (b) Average bias for a 3 component model of fat in cheese predicted from replicate hyperspectral images. Data pre-treatments were: none (star), first derivative (circle), common wavelength range absorbance (square), and common wavelengths + first derivative (triangle).

replicate measurements of the LOO cheese were removed from the calibration set as well, to avoid any model alias effects. As with the artificial mixtures, a region of interest mask was used to select a subset of spectra from the hyperspectral images, from which median spectra were computed. An *average cheese* was selected to see how well models built on the remaining cheese types would perform. Independent calibration models and test set predictions were made from five sets of HI images, one set of SG spectra, and three sets of FO spectra with replicates. The expected values for this *average cheese* were 30% protein, 17% fat, and 1.5% carbohydrate.

Fig. 3(b) summarizes the predictive results for fat obtained from five independent sets of replicate samples measured with the HI instrument. Results from four types of spectral preprocessing are presented: the full wavelength absorbance spectra, first derivative spectra, and absorbance and first derivative spectra but limited to the 1100–1662 nm wavelength range common between instruments. Once again the major variance observed in this and similar plots with results of other samples or other instruments was that associated with sampling and instrument errors. No clear trends were

**Table 4** Bias and standard deviations of replicate measurements of one cheese. Calibration models were based on all other cheeses, with values provided by manufactures, or (*) from laboratory analysis

| | Protein | | Fat | | Carbohydrate | | Protein* | | Fat* | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std | Bias | Std | Bias | Std | Bias | Std | Bias | Std |
| Absorbance | | | | | | | | | | |
| HI | −2.0 | 2.8 | 1.9 | 3.3 | 0.5 | 0.7 | −1.3 | 3.5 | 0.1 | 0.4 |
| SG | −3.0 | | 0.2 | | 0.7 | | −1.6 | | 1.2 | |
| FO | −1.4 | 1.0 | −1.0 | 1.4 | −0.1 | 0.7 | −1.6 | 1.3 | −0.9 | 1.3 |
| First derivative | | | | | | | | | | |
| HI | −1.5 | 2.3 | −0.9 | 0.9 | −0.1 | 1.1 | −0.6 | 0.5 | −0.2 | 1.3 |
| SG | −0.8 | | −1.6 | | 0.1 | | 0.2 | | −0.5 | |
| FO | −1.9 | 2.0 | −1.2 | 1.2 | 0.4 | 0.9 | −3.0 | 2.7 | −0.7 | 0.8 |

observed in the effects of the spectral pre-treatments, or between instruments.

In addition to the protein, fat, and carbohydrate content reported by the cheese manufacturers, the cheeses were analyzed independently for protein and fat content. The concentration values varied slightly, and were used for independent modeling. Prediction results for all constituents with all instruments are listed in Table 4. General bias and standard deviation values were about the same as those obtained in the artificial mixture analysis. The overall bias results of the three instruments were quite similar, however the observed precision decreased significantly from SG to HI to FO instruments. All the instruments used were based on diffuse reflection. The SG instrument scanned over a large area, and gave very precise results. The HI was not as precise as the SG instrument, but because of the averaging effect of thousands of sample spectra it did very well. The FO instrument provided the highest wavelength resolution and was considered very stable, but the fiber optic probe introduced additional uncertainty. The major variation observed with all instruments seemed to be due to sample preparation and presentation technique.

The lab results for protein and fat were for the actual cheeses and were considered reliable. The values reported on package labels agreed well, but were less reliable because they were only for typical batches. Most HI results were therefore good. They were within the error expected between the label value and actual cheese value. The agreement with the lab results was also good.

### Hyperspectral image advantages

The results discussed so far have been based on the comparison of a single median spectrum from each hyperspectral image with the spectra obtained from the physical averaging obtained with the other spectrometers. The main advantage of the hyperspectral image is that it contains over 81 000 spectra from which large sample population statistics and spatial information can be extracted. The calibration models previously computed from median spectra were applied to the sets of spectra defined by the spatial ROI selection masks, resulting in thousands of prediction values. These values can be summarized in a univariate histogram, or depicted as a concentration map by assigning each prediction value to the spatial coordinates provided by the sample selection mask. Fig. 4 shows these results for the 4 component PLS model predictions of citric acid from one of the center point mixture images. This figure provided very significant information not available by the median spectrum prediction or the spectra acquired from the SG or FO spectrometers. As with the other instruments, the mean of the distribution was very close to the expected value, 34.3 percent, but there was an extremely wide distribution of predicted values ranging between 0 and over 60 percent. The spatial concentration map indicated how this non-uniformity was distributed throughout the measured sample. The lower half of Fig. 4 represents the predictions of the same masked area, but based on first derivative spectra. Again the mean value of the prediction distribution was very close to the expected value. The peak width was substantially narrower, and the intensity contrasts in the spatial concentration map much less. This suggested that a significant baseline offset type correction was being made by the first derivative transform.

Additional interpretation of hyperspectral image prediction histograms and spatial maps can be gained by using purposely overfit and underfit PLS calibration models. As additional latent variables are included, signal variance is included which may or may not correlate to the independent $y$ block variables. However when this information is viewed as a spatial map, information regarding sample surface defects or illumination inconsistencies may appear. Such sample and instrument diagnostic information is not easily available from the non-imaging spectrometric techniques.

Fig. 5 shows similar results for the protein analysis of one of the cheeses. The relative shape and offset of the two histograms may assist in determining the correct number of components, or best preprocessing technique to use. In this case the two spatial maps help identify two sample problems. Cutting marks are visible as streaks in the upper image, but were removed by the first derivative transform. This is a surface effect that contributes to the accuracy of the prediction results. The first derivative image also showed areas of much higher protein concentrations. This may indicate non-uniform distribution of the protein, or possibly be an artifact caused by drying of the surface from the high intensity illumination lamps. But in either case, this is additional information not available with the SG and FO spectrometers.

It is essential to interactively explore the different spectral and spatial aspects of the hyperspectral images. Additional software was developed to facilitate exploratory analysis using PCA decompositions. This software allows the simultaneous presentation of images and two-dimensional score plots. The PCA scores for every pixel can be presented as images, compressing the information contained in tens or hundreds of image channels into a very few significant score images. Scatter plots of score pairs obtained from two principal components can also be made, permitting visualization of the clustering of samples with similar spectral characteristics. Because there are so many sample points often overlapping in these score plots, it is advantageous to additionally color-map the 2D plots to indicate the relative level of counts of each pair of score values. From these sets of score images and score plots, single points or sets of points contained in ROIs may be selected in one plot, and mapped into all other plots. Fig. 6 is one example of this

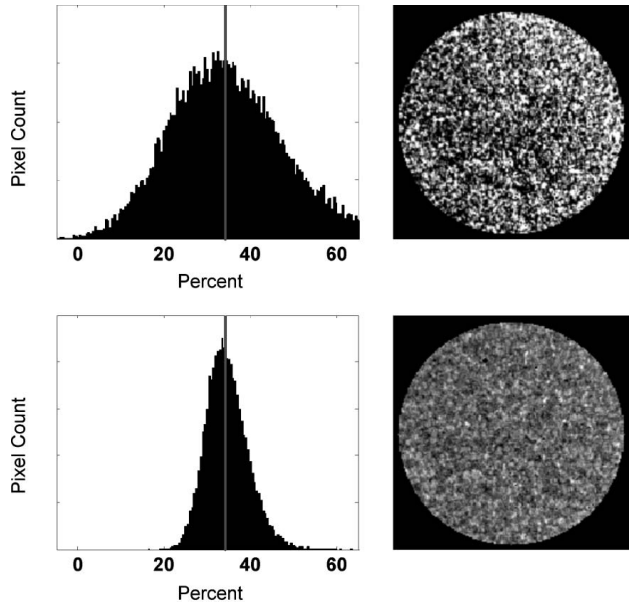


**Fig. 4** Center point mixture predictions for citric acid concentration (expected concentration: 33.4 %) for a 4 component model. Histograms and concentration maps are for absorbance spectra (top) and first derivative spectra (bottom). 20 000 sample spectra were selected using a circular ROI mask with an 80 pixel radius.

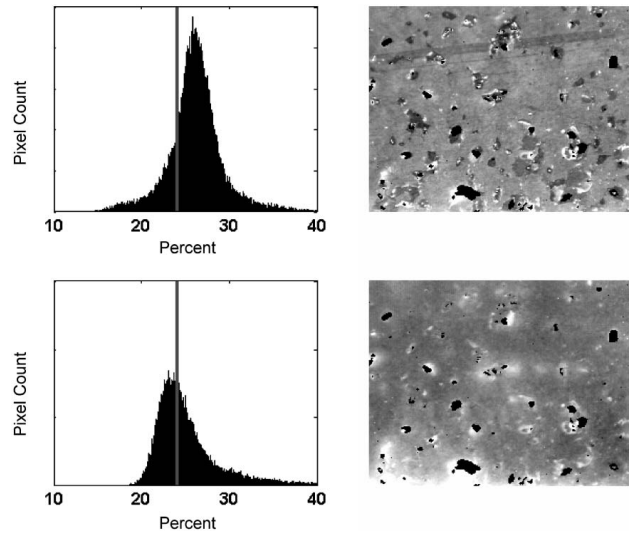


**Fig. 5** Protein concentration predictions (expected concentration: 24 %) for a 4 component model. Histograms and concentration maps are for absorbance spectra (top) and first derivative spectra (bottom). ~48 000 sample spectra were selected using a rectangular mask, 195 × 250 pixels.

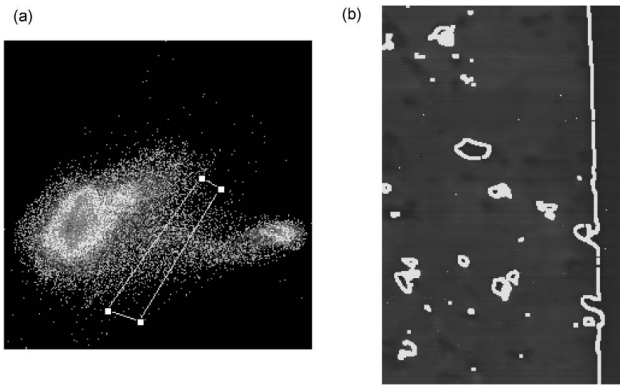


**Fig. 6** Interactive PCA exploration of cheese. Score pairs are first scaled to map all samples into a 256 × 256 2D histogram scoreplot (a). The spatial coordinates of all samples selected within the ROI of (a) are then mapped onto score image (b), in this case, indicating shadow or edge effects of cheese holes.

process. A region of interest in the $t_2$–$t_5$ score plot (Fig. 6(a)) has been selected. The pixels associated with these score values were then identified and marked in all other image and score plots currently displayed. In this case, a class of pixels having to do with shadows or edges of holes in the cheese has been identified, which has been indicated by the white masking in Fig. 6(b).

This interactive exploration of both spatial and spectral information is unique to hyperspectral image data, and can be performed in either direction: score space points selected and mapped into image space, or image pixels selected and mapped into score space. Specific pixels selected in this manner can then be used as mask regions for use in obtaining more robust PLS prediction models.

## Conclusion

Hyperspectral images offer a natural expansion of conventional spectroscopy, adding spatial positioning information to the sets of spectra acquired. The resulting data-rich hypercube must be carefully examined to fully extract the information hidden within. Simple spatial compression techniques reducing the entire hypercube or a subset region of interest to a single average spectrum show comparable global predictive ability with other spectrometers. Calibration models based on these average ROI spectra can also be used to predict constituent concentrations at each individual pixel in new test set images. Accurate large sample population statistics and prediction value histograms can then be generated, providing information regarding the uniformity of constituent concentration. Mapping the concentration values to pixel locations adds spatial information allowing even greater insight into sample uniformity.

In the two datasets examined, the use of data preprocessing transforms did little to change the average overall predicted concentration value of constituents. Just as much variance was observed from sample replicates as from altering the preprocessing treatment. However the first derivative transform provided a major change in the range in values and spatial distribution of the prediction results obtained from applying the calibration models to the individual spectra in sample regions of interest. Additional transforms had minimal impact and were overshadowed by sample handling and instrumentation variations. Sample selection and preparation is a critical part of obtaining accurate quantitative results and needs to be examined further.

Hyperspectral NIR imaging is a hybridization of spectroscopy and spatial imaging. The decrease in spectral wavelength resolution currently available in NIR imaging systems is offset by the increase in spectral quality obtained from averaging thousands of spectra. Although lengthy image acquisition times may be a problem for some biological samples which either move or are temperature sensitive, the overall quantitative results obtained are comparable to those of spot probe spectroscopic techniques. The spatial concentration mapping and sample uniformity information provided by hyperspectral imaging is a benefit not easily obtainable with other spot probe spectroscopic techniques. Specialized software which promotes the simultaneous exploration of both spatial and spectral information enables further data-mining not attainable with other spectroscopic techniques.

## Acknowledgements

The authors thank NIRCE, an EU Unizon-Kwarken project and SLUP for financial support of this project. We also acknowledge Carina Jonsson, SLU, for performing additional cheese fat and protein analysis. Kempestiftelserna grant SMK-2062 provided funding for instrumentation equipment.

## References

1 B. Osborne, T. Fearn and P. Hindle, *Practical NIR Spectroscopy*, Longman Scientific and Technical, Essex, 2nd edn, 1993, ch. 8, pp. 145–199.
2 R. Nortvedt, O. Torrissen and S. Tuene, *Chemom. Intell. Lab. Syst.*, 1998, **42**, 199–207.

 

3 I. Masserschmidt, C. Cuelbas, R. Poppi, J. De Andrade, C. De Abreu and C. Davanxo, *J. Chemom.*, 1999, **13**, 265–273.
4 O. Berntsson, L. Danielsson, B. Lagerholm and S. Folestad, *Powder Technol.*, 2002, **123**, 185–193.
5 T. Lestander and P. Geladi, *Analyst*, 2003, **128**, 389–396.
6 G. Hodge and W. Woodbridge, *J. Near Infrared Spectrosc.*, 2004, **12**, 381–390.
7 Spectral Dimensions, Inc., 3416 Olandwood Court, Suite 210, Olney, MD 20832, USA.
8 P. Geladi, S. Wold and K. Esbensen, *Anal. Chim. Acta*, 1986, **191**, 473–480.
9 K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1989, **7**, 67–86.
10 P. Geladi and H. Grahn, *Multivariate Image Analysis*, Wiley, New York, 1996.
11 P. Geladi and K. Esbensen, *J. Chemom.*, 1991, **5**, 97–111.
12 K. Esbensen, P. Geladi and H. Grahn, *Chemom. Intell. Lab. Syst.*, 1992, **14**, 357–374.
13 T. Lied, P. Geladi and K. Esbensen, *J. Chemom.*, 2000, **14**, 585–598.
14 Y. Honglu and J. MacGragor, *Chemom. Intell. Lab. Syst.*, 2003, **67**, 125–144.
15 J. Burger and P. Geladi, *J. Chemom.*, 2005, **19**, 355–363.
16 Foss NIRSystems, Slangerupgade 69, Postbox 260, DK-3400 Hillerød, Denmark.
17 Bruker Optics Inc, 19 Fortune Drive, Manning Park, Billerica, MA 01821–3991.
18 ImageJ, *Image Processing and Analysis in Java*, http://rsb.info.nih.gov/ij/.
19 JAMA basic linear algebra package for Java, National Institute of Standards and Technology, http://math.nist.gov/javanumerics/jama/.
20 JFree software projects, http://www.jfree.org/jfreechart/index.php.

# IV

# Spectral pre-treatments of hyperspectral NIR images: analysis of diffuse reflectance scattering

**James Burger* and Paul Geladi**

Unit of Biomass Technology and Chemistry, Swedish University of Agricultural Sciences, PO Box 4097, SE 90403 Umeå, Sweden

email: jim.burger@btk.slu.se voice: +46 90 786 8798 fax: +46 90 786 8799

Scattering effects are often encountered when measuring diffuse reflectance near infrared (NIR) spectra of solid and semi-solid materials. How does this phenomenon effect hyperspectral imaging of powders? A series of hyperspectral NIR images of particle size fractions of commercial grade salt and sugar were acquired. Spectral pre-processing techniques including Kubelka-Munk, standard normal variate, and absorbance transforms, unit length or unit area normalization, first and second derivative transforms, and several variants of multiplicative scatter corrections (MSC) were applied to the images and examined for their effectiveness at reducing or eliminating scatter effects. Principal component analysis (PCA) scoreplots produced expected results: derivative transforms reduced variance, but did not eliminate the particle size dependencies; piecewise MSC transforms reduced the data to two clusters, one for salt and one for sugar. Partial least squares (PLS) regression was applied to examine the impact of the pre-processing transforms on prediction of particle size. RMSEP values between 10 and 50 micrometers were determined for particle fractions ranging between 140 and 315 micrometers for all transforms except the piecewise MSC; in spite of the reduction in additive and multiplicative effects, enough correlated variance remained after application of the pre-processing transforms to allow prediction of particle size ranges from PLS models. Additional scatter effect information was obtained by examining particle size distribution histograms and spatial particle size mappings facilitated by the hyperspectral images.

*Keywords:* Hyperspectral image, scatter correction, pre-treatment, near-infrared

---

## Introduction

Much has been written about various numerical preprocessing treatments proposed to linearize spectral response and reduce or eliminate additive or multiplicative effects in near infrared (NIR) spectra.[1-4] The most basic approach is to simply convert the diffuse reflectance *R* measurements into more linear units such as absorbance which can then be directly related to concentration. Alternatively the Kubelka-Munk transform can be applied.[5] When changing baseline offsets are encountered, first and second derivative transforms may be applied to remove the offset.[6,7] This is typically done in spectroscopy using standard Savitsky-Golay polynomial filters.[8-10] To correct for multiplicative effects such as those introduced by light scattering from different sized particles, multiplicative scatter corrections[11-17] (MSC), piecewise MSC[18] (PMSC), and extended MSC have been successfully implemented.[19] The standard normal variate (SNV) transform[20] has also been used, however this can be shown to be linked to the MSC transform.[21]

The spectral perturbations encountered in NIR spectroscopy addressed by this collection of pre-processing treatments occur in the spectra contained in hyperspectral NIR images as well, although the spectroscopic geometry of an integrating spectrometer and that of an imaging camera are quite different.

How do the spectral pre-treatment transformation techniques affect the quality of a hyperspectral image containing tens of thousands of spectra? How do light scattering effects from different sized particles propagate in the spectra acquired at adjacent pixel locations within a hyperspectral image? The purpose of this work is to assess the impact of pre-processing treatments on a series of hyperspectral images obtained from two sets of particle size fractions of commercial grade sugar and salt. The ability of these treatments to negate the effect of the variations in the particle sizes will be examined

## Experimental

### Samples

Two sets of samples were prepared from consumer grade sugar (Pärlsocker, Danisco Sugar AB) and salt (Grovt Salt, Ab Hanson & Möring). Course particles were placed in a kitchen food processor (OBH Nordica) and chopped for 1 minute. The mixture of particle sizes was then separated into nine size fractions using a series of woven wire mesh sieves (Endecotts Ltd.) with screen openings ranging from 600 to 63 micrometers using a mechanical shaker system for 10 minutes. The individual fractions retained a significant amount of fines, and were therefore re-separated one at a time, placing each fraction at the top of a set of three sieves, starting with a mesh size two sizes larger than the final targeted fraction size end point. This final separation was done by manual shaking for larger particles, and with the aid of the mechanical shaker for the very small particles. Table 1 indicates the resulting nine size fractions obtained for both the sugar and salt particles.

Table 1 Particle Size Distribution Limits and Ranges

| Sample Fraction | Size Range (µm) |
|---|---|
| A | > 400 |
| B | 315 – 400 |
| C | 250 – 315 |
| D | 200 – 250 |
| E | 140 – 200 |
| F | 125 – 140 |
| G | 100 – 125 |
| H | 063 – 100 |
| I | < 063 |

Approximately 6 grams of each fraction was placed in a circular NIR sample holder (Foss NIRSystems) and measured with a Foss NIRSystems 5000 scanning grating monochromator equipped with a rotating sample holder. The wavelength range selected was 1100-2498 nm with 2 nm resolution. Spectra were mean centered and processed with principal component analysis (PCA.) The resulting $t_1$ – $t_2$ score plots were examined for general trends in spectral deviations related to particle size. Several samples appeared out of line, and were hand sieved an additional time, producing a set of samples having score plots with consistent particle size trends for both sugar and salt fractions.

**Instrumentation**

Hyperspectral NIR images or hypercubes were acquired using a MatrixNIR imaging system (Spectral Dimensions, Inc.) The spectral range used was 960 to 1662 nm with 6 nm resolution. Ten replicate scans were acquired and averaged from each sample, producing hyperspectral images or hypercubes with 256 x 320 pixels and 118 spectral channels. The entire camera field of view was first calibrated for reflectance by measuring a series of five Spectralon standard reference material tiles with 99, 75, 50, 25, and 2 percent reflectance. This set of reflectance standards was measured at the beginning and end of every set of samples. The Foss NIR sample holders were also used for imaging purposes, providing a flat sample surface necessary for the limited focal distance of the MatrixNIR optics. The sample height for each sample was adjusted to maximize image sharpness. The 60 x 52 mm image field of view (200 x 200 micrometer pixels) also included a gray scale series of Spectralon standard reference materials for reflectance calibration, as well as a rectangular piece of NIST glass standard reference 2035A for wavelength calibration. The use of these materials for instrument standardization and system diagnostics as well as the larger standards for reflectance calibration has been described in previous papers.[22, 23] Figure 1 shows a typical image from one channel measured at 1074 nm. The circular sample holder, the Spectralon gray scale, and the rectangular glass standard can be easily seen. Duplicate samples were taken from all sample fractions, producing a second set of test images. A third set of replicate images was also measured, using a higher magnification lens. This last series of images had a field of view of 10 x 12 mm (40 x 40 micrometer pixels) which was completely filled with sample. The internal standards were outside the field of view for this magnification.

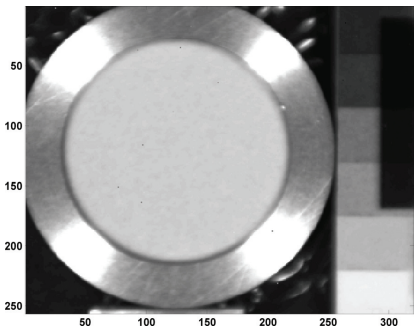


Figure 1. A single channel image (256 x 320 pixels at 1074 nm) including circular sample holder and internal standard reference materials for monitoring reflectance (gray scale) and wavelength (dark rectangle) stability. The circular sample region contains ~18000 usable pixels

**Numerical processing**

*Reflectance transforms*

All hyperspectral images acquired were stored in raw data format with data points ranging between 0 and 4095, the range of the 12 bit A/D converter. The series of five reference Spectralon hypercubes permits the computation of individual second order reflectance transform functions for every pixel location at every wavelength channel. These transform functions were applied to each sample hypercube to produce hyperspectral images transformed to diffuse reflectance units. Although each sample image contained internal standards, further corrections for instrument standardization based on the image internal standards were not done. However spectra from the internal standards in all images were compared to confirm stability of the instrument.

*Pixel selection and outlier detection*

Because of the geometry of the Foss sample holders, slight shadow effects were observed in the perimeter area within the circular sample region. A slightly smaller region with a radius of 80 pixels was chosen to eliminate this effect, resulting in a sample pixel count of ~20000 points. The detection device used in this imaging system also contains *dead* pixels which produce faulty data or outlier points. An iterative two step process was employed to detect and remove resulting outlier spectra. Since each image contained in theory thousands of spectra belonging to a single sample class $\boldsymbol{X}$, (particle size distribution), outliers from this class were detected based on estimates of class population statistics. First the mean population spectrum was computed as a target spectrum, $\boldsymbol{x}$. Two metrics were then computed for all spectra: $d$, the Euclidean distance to $\boldsymbol{x}$, and $c$, the cosine of the angle to $\boldsymbol{x}$. The mean and standard deviations of $\boldsymbol{d}$ and $\boldsymbol{c}$ were used to selectively trim away outliers, spectra whose metrics exceeded a three standard deviations displacement. Both the mean and standard deviation are heavily influenced by outliers, so after trimming any outliers a new target spectrum $\boldsymbol{x}$ was computed and the outlier detection process repeated. It was found empirically that neither metric test alone was sufficient; angle outliers remained after removing all distance outliers, and vice versa. It was also found that after applying this cleaning process to a set of reflectance

spectra, followed by a spectral pre-treatment, additional outliers could be detected. Consequently the iterative outlier detection and removal process was repeated after the application of any spectral preprocessing treatment. Of course all outliers should be examined before discarding, however there is a general advantage to hyperspectral images: because there is such an abundance of spectra, a rather liberal outlier detection scheme can be applied and still allow the retention of a very large number of sample spectra. For the discussion that follows, it is assumed that all image sample spectra sets have been thoroughly cleaned, that is all outlier spectra have been detected and removed. This typically left ~18000 to ~19000 sample spectra from each hyperspectral image. Sample mean spectrum and population statistics were based on these fully cleaned datasets, however in some cases for computational efficiency these large spectra sets were further reduced in size by sample subselection; only every $k^{th}$ spectrum was selected for analysis.

*Additional computations*

In addition to the spectral pre-treatments, principal component analysis (PCA) and partial least squares regression (PLS) were performed. All computations were based on the entire spectral range of the acquired spectra. No variable selection was done. For the purposes of PLS, the spectra are considered predictor variables or X block data. Response or Y block data are related variables such as concentration or other measurable physical properties. Computations were performed using MATLAB 6.1, using scripting code written by the authors.

## Results and discussion

**Mean image spectra**

Mean spectra were computed form the cleaned spectra sets from 18 sample images: 9 sugar and 9 salt fractions. Two observations can be made from the raw absorbance spectra displayed in Figure 2a. Firstly, salt has a very weak absorbance in the NIR, whereas sugar has a relatively large absorbance especially above 1400 nm. Secondly, as expected from light scattering effects, a systematic decrease in absorbance was observed with a decrease in particle size. This was evident in both the sugar sample spectra, and also the weaker salt sample spectra: the greatest absorbance was measured in the largest particle size fractions. Figure 2b shows the effect of the first derivative transform. Baseline offsets have been removed however particle size dependencies are still clearly evident. A global piecewise MSC (GPMSC) was performed choosing the mean of all sugar or salt spectra as a target value. The GPMSC transform collapses all spectra towards a common target spectrum, clearly seen in Figure 2c where the corrected spectra form two narrow bands. The corrected spectra for both salt and sugar are very similar below 1370 nm, but are distinguishable above this wavelength.

**Principal component analysis**

The spectra displayed in Figure 2 represent the results from first compressing the entire hyperspectral image into a single mean spectrum from each sample, effectively producing the same results as would be obtained from a typical integrating NIR instrument. PCA was employed to more closely examine the variations within the abundance of sample spectra contained in each hyperspectral image. An augmented spectral matrix was created by appending the ~20000 spectra from 16 sample images (both salt and sugar) and subjected to PCA. For display and computational purposes, only every $80^{th}$ spectrum was included, resulting in a matrix of ~4000 spectra. Additionally since the smallest particle size fractions for both salt and sugar were extremely noisy, they were excluded from this analysis because of their significant variance contribution. Figure 3 shows the mean centered PCA $t_1$ – $t_2$ score plots for the raw absorbance spectra before and after spectral pre-treatments. The points from each particle size fraction have been color coded (gray scale) to help discriminate between particle sizes. An arrow has been included to indicate the relative direction from small to large particle size fractions. All sugar spectra are represented with a star symbol, and all salt spectra with a small circle. These plots reveal many interesting spectral features uniquely available from hyperspectral images.

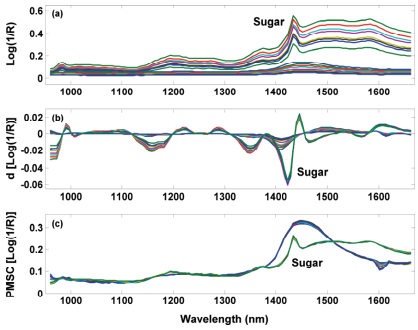


Figure 2. Mean image spectra for 8 sugar and 8 salt particle size fractions: raw absorbance spectra (a), first derivative spectra (b), and global (salt or sugar set) piecewise MSC spectra (c). Raw absorbance is greater for sugar, and decreases with particle size.

Figure 3a represents the clustering of raw absorbance spectra. The variance along the $t_1$ axis is significantly greater for sugar, as would be expected from the greater variance in absorbance observed in the sugar spectra presented in Figure 2a. However the variance on the $t_2$ axis is nearly the same for sugar and salt. In both cases a systematic trend can be seen with respect to particle sizes. A trend in the amount of variance within each size fraction is also visible in the sugar clusters: as the particle size increases, so does the size of the score plot cluster due to the variance in the individual spectra. The measured light scattering effect has a greater variance with larger particle sizes. This is possibly due to a sample averaging effect: for any given pixel area, there is a greater number of small particles than large. This imposes a statistical averaging which reduces the measured variance of the small particles relative to the large

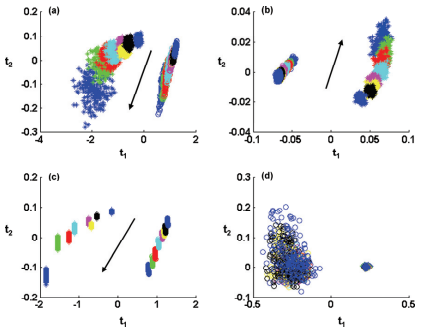


Figure 3. PCA scoreplots for ~250 spectra from each particle fraction image, salt (circle) and sugar (star). Scoreplots are based on raw absorbance spectra (a), first derivative spectra (b), MSC applied to each individual set of image spectra (c), and global piecewise MSC applied to all salt or sugar spectra (d). An arrow indicates the relative direction from small to large particle size fractions.

The effect of baseline offset corrections can be seen in the score plot of Figure 3b for the first derivative transform: the clustering of points within each individual image is much tighter. The particle size dependency is however still clearly visible. A similar plot of second derivative scores revealed a similar clustering structure for sugar, while the size dependency for salt was nearly removed. MSC corrections on an image by image basis further reduced the variance within each image, as indicated in the clustering in Figure 3c. With this pre-treatment, the mean of the spectra from each image should be nearly identical before and after MSC correction, hence the over all class structure of Figures 3a and 3c were very similar. The global piecewise MSC correction ideally projects all spectra towards a single target spectrum, in this case one spectrum for sugar and one for salt. Figure 3d indicates the result of this correction treatment. The entire set of sugar spectra now forms a very small cluster. The particle size dependency has been removed from the salt spectra as well, but because of greater relative noise levels, these points form a significantly larger sized cluster. Nonetheless, it appears that this pre-treatment removes the particle size component from the image spectra.

**Partial least squares predictions**

*Calibration / test set selections*

One of the primary goals of spectral pre-treatments is to linearize or correct spectra to improve regression modeling and predictions of analyte properties. Since the measured hyperspectral images showed a strong correlation with particle size information, the performance of PLS models to predict the sugar particle sizes, before and after different spectral pre-treatments was investigated. What characterizes a particle size distribution? The true size distribution of each fraction was not known. The particles studied were not monodisperse, they formed a distribution achieved by an imperfect sieving process. If the actual particle size distribution had been known, properties such as mean or median particle size, or distribution moments such as width or skewness could be expressed. This was not the case with the experimental sample fractions, so some parameter obtained from the sieving process had to be used as a distribution descriptor. The lower size endpoint of each size fraction was chosen as an external reference value. This value was not known for the smallest size particle fraction, so the smallest size fraction was eliminated from modeling and prediction. The mean spectrum from the clean set of spectra from each calibration set image was used for modeling purposes, resulting in PLS models computed from 8 spectra with 118 variables. In all cases a one component PLS model accounted for nearly all predictor and response variable variance.

Predictions were made from the complete set of clean spectra from the same set of images used for calibration model creation, a second set of independent test images and a third set of test images acquired under higher magnification. Prediction results were also compared with and without fraction size cross validation; that is, using all 8 sample size fractions in the PLS model, vs. removing the spectrum from the calibration set that corresponded to the fraction used for testing. The results from all of these calibration / test set scenarios were nearly identical. The results presented here are based on the mean spectra from one set of images used for calibration, and the second set of independent test set image spectra used for predictions, typically 18000 – 19000 spectra from each image. All 8 fraction sizes were included in the calibration model.

*Summary prediction statistics*

Figure 4 displays the RMSEP values for different pre-treatments of the different particle size fractions. The individual image based MSC contributed to a significant decrease in RMSEP for the largest particle size fractions, where spectral variation was greatest. The global based MSC reduced the variance between the spectra used for the calibration model, and consequently the RMSEP values for all particle sizes were higher for this pre-treatment. Otherwise all pre-treatments resulted in quite similar RMSEP values, with a slight decrease in values associated with the derivative pre-treatments. In spite of the changes in clustering exhibited in the scatter plots of Figure 3, PLS does a very good job at finding any correlated variance between predictor and response variable blocks.

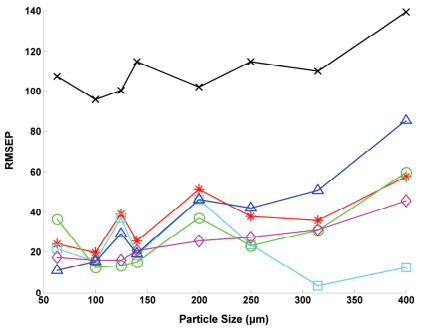


Figure 4 Particle size prediction RMSEP values for ~18000 spectra from images of each sugar size fraction. Results are for raw absorbance (star), and individual image based MSC (square), global piecewise MSC (cross), first derivative (circle), second derivative (diamond), or Kubelka-Munk (triangle) pre-processing transforms.

*Hyperspectral image results*

The RMSEP value represents a single statistical value characterizing each of the pre-treatment / size fraction combinations. Hyperspectral imaging permits other representations of the same prediction results. Histograms of the over 18000 predictions from a single image can be examined and provide an indication of size distributions. Figure 5 displays the histograms of the one component PLS predictions for the B fraction of sugar, with an expected particle size of 315 µm indicated by the vertical line. The distribution of raw absorbance based predictions (histogram a) is well centered on this expected value however the predicted values range from 200 – 400 µm. The image based MSC yields a very narrow distribution (histogram b); the derivative treatments narrow the distributions slightly (histograms d and e); and finally the global MSC and Kubelka-Munk treatments broaden the distributions (histograms c and f.) These general trends in predicted size distribution resulting from the various spectral pre-treatments were observed for all samples and are not identifiable in the simple RMSEP values of Figure 4.

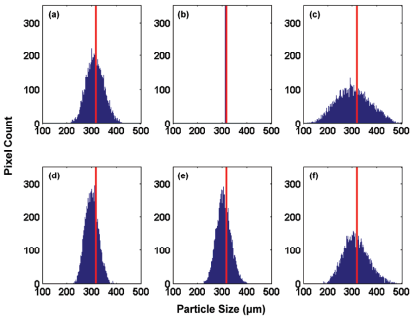


Figure 5. Distributions of ~ 18000 particle size predictions for the 315 µm sugar fraction based on one component PLS models of raw absorbance (a), and individual image based MSC (b), global piecewise MSC (c), first derivative (d), second derivative (e), or Kubelka-Munk (f) pre-processing transforms. Vertical lines indicate target 315 µm size.

The same sets of individual spectra predictions can also be mapped into their spatial image locations, producing 2-dimensional prediction maps as in Figure 6. In this case the gray scale images have been mapped such that black to white represents the particle size range 150 – 500 µm. The background squares surrounding each of the circular prediction maps has been set to the target value of 315 µm. This representation provides another perspective for interpretation of prediction results. Relative to the raw absorbance based predictions (image a), the very narrow distribution of image based MSC predictions appears as a very flat prediction image (image b); the broader prediction distributions of the global MSC and Kubelka-Munk treatments appear as course texture (images c and f); while the derivative based predictions appear slightly smoother in texture (images d and e.) As with the particle size distribution histograms presented in Figure 5, these general results were obtained for all samples, offering additional size distribution scattering effect details not achievable with integrating NIR spectrometers.

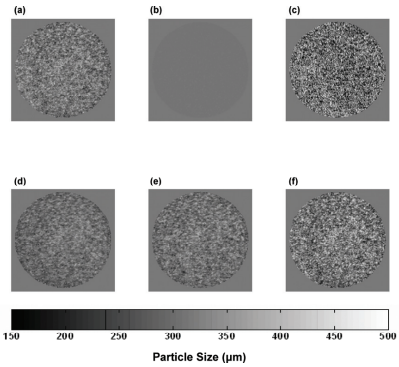


Figure 6. Prediction image maps of ~ 18000 particle size predictions for the 315 µm sugar fraction based on one component PLS models of raw absorbance (a), and individual image based MSC (b), global piecewise MSC (c), first derivative (d), second derivative (e), or Kubelka-Munk (f) pre-processing transforms.

*Mixture analysis*

The prediction results obtained suggest that PLS models can be constructed to determine particle sizes with reasonable accuracy. To test this further, a mixture was made of equal amounts of material from sugar fractions B and H, having size distributions of 315 – 400 and 063 – 100 µm respectively. Unfortunately particle scattering effects do not add the way the absorbance of different chemical constituents do. Figure 7 shows the distributions for the size predictions of the second derivative spectra of the B, H, and BH mixture samples. The narrow size fraction

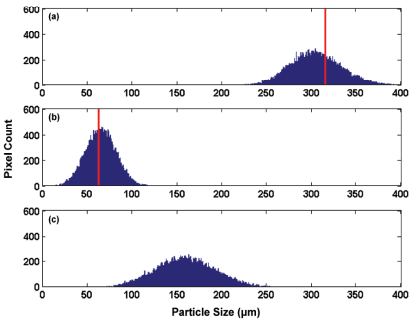


Figure 7. One component PLS model based particle size prediction distributions for ~ 18000 second derivative sugar spectra each from images of the 315 µm (a), and 63 µm (b) fractions, and a 1:1 mixture of the same two fractions (c). Vertical lines indicate target sizes of 315 µm and 63 µm.

samples B and H are predicted quite well, however the combination of B and H samples does not yield a bi-model prediction distribution. In this case, the smaller particles of sample H have filled in some of the holes of sample B, producing light scattering effects with an intermediate nature.

## Conclusion

The spectral pre-treatments of the mean spectra from hyperspectral NIR images give similar results to those obtained with traditional integrating NIR instruments. However examination of the thousands of NIR spectra within each hyperspectral image provides additional information not available from these classical spectroscopic techniques: PCA score plots and histograms and spatial mappings of PLS predictions provide insight into the bulk and spatial distributions of spectral deviations and their dependencies on the results of pre-processing treatments. Derivative transforms reduce the within image variance (additive effects), but not the particle size dependencies (multiplicative effects). The GPMSC transform performs the best at minimizing both these effects. These results are not unexpected however it was surprising to find that PLS could still find enough correlated variance to enable particle size prediction. No prior knowledge was known regarding the true particle size distributions in the samples measured; a similar analysis of monodispersive powders or particles with known particle size distributions should be examined to further examine the effects of spectral pre-treatments on hyperspectral images.

## Acknowledgements

The authors thank SLUP and NIRCE, an EU Unizon-Kwarken project for financial support of this project. Additionally Kempestiftelserna grant SMK-2062 provided funding for instrumentation equipment. Also thanks to Steven Choquette at NIST for providing the glass standard reference material.

## References

1. T. Naes, T. Isaksson, T. Fearn and T. Davies, *A User Friendly Guide to Multivariate Calibration and Classification*. NIR Publications, Chichester, UK (2002).
2. S.R. Delwiche and J.B. Reeves, *J. Near Infrared Spectrosc*. **12**, 177-182 (2004).
3. V. Svedas, *J. Near Infrared Spectrosc*. **12**, 347-358 (2004).
4. B.G. Osborne, T. Fearn and P.H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific & Technical, Essex, UK p. 99 – 119 (1993).
5. P. Kubelka and F. Munk, *Z. technische Physik*, **12**, 593-604 (1931).
6. K.H. Norris and P.C. Williams, *Cereal Chem*. **61**, 158 (1984).
7. F.G. Giesbrecht, W.F. McClure and A. Hamid, *Appl. Spectrosc*. **35**, 210 (1981).
8. A. Savitzky and M. Golay, *Anal. Chem.* 1964; **36**: 1627-1639.
9. J. Steiner, Y. Termonia and J. Deltour, *Anal. Chem.* 1972; **44**: 1906-1909.
10. H. Madden, *Anal. Chem.* 1978; **50**: 1383-1386.
11. H. Martens, S.A. Jensen and P. Geladi, in *Proceedings of the Nordic Symposium in Applied Statistics*, Ed by O.H.j. Christie Stokkand Forlag, Stavanger, p. 205-234 (1983).
12. P. Geladi, D. McDougall and H. Martens, *Appl. Spectrosc*. **39**, 491-500 (1985).
13. I.S. Helland, T. Naes and T. Isaksson, *Chem. Lab*. **29**, 233-241 (1995).
14. A.J. Iversen and T. Palm, *Appl. Spectrosc*. **39**, 641-646 (1985).
15. J.L. Ilari, H. Martens and T. Isaksson, *Appl. Spectrosc*. **42**, 722-728 (1988).
16. T. Isaksson and T. Naes, *Appl. Spectrosc*. **42**, 1273 – 1284 (1988).
17. T. Naes, T. Isaksson and B.R. Kowalski, *Anal. Chem*. **62**, 664-673 (1990).
18. T. Isaksson and B.R. Kowalski, *Appl. Spectrosc*. **47**, 702 – 709 (1993).
19. H. Martens and E. Stark, *J. Pharm. Biomed. Anal*. **9**, 625-635 (1991).
20. R.J. Barnes, M.S. Dhanoa and S.J. Lister, *Appl. Spectrosc*. **43**, 772-777 (1989).
21. M.S. Dhanoa, S.J. Lister, R. Sanderson and R.J. Barnes, *J. Near Infrared Spectrosc*. **2**, 43-47 (1994).
22. J. Burger and P. Geladi, *J. Chemometr*. **19**, 355-363 (2005).
23. J. Burger and P. Geladi, *J. Chemometr*. (accepted) (2006).