2013

Calibration optimization and efficiency in near infrared spectroscopy

Nanning Cao
Iowa State University

Recommended Citation

Cao, Nanning, "Calibration optimization and efficiency in near infrared spectroscopy" (2013). *Graduate Theses and Dissertations*. 13199.
http://lib.dr.iastate.edu/etd/13199

Calibration optimization and efficiency in near infrared spectroscopy

by

Nanning Cao

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agriculture Engineering

Program of Study Committee:
Charles R. Hurburgh Jr., Major Professor
Carl J. Bern
Chenxu Yu
Dianne H. Cook
Jay-Lin Jane

Iowa State University

Ames, Iowa

2013

This dissertation is dedicated to

my grandfather, the late Huifu Shi (石惠甫) , who taught me to face life cheerfully

and

my uncle, the late Taoyuan Shi (石桃园), who encouraged me to pursue my dream of future

that has lightened my way

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First I would like to express my gratitude and regards to my major professor Dr. Charles R. Hurburgh Jr. for his support and helpful guidance, which helped me to complete my piece of work through various stages. I highly appreciate the valuable contribution of staff and graduate students in Grain Quality Laboratory. I would like to thank member of my program of study committee, Dr. Carl J. Bern, Dr. Chenxu Yu, Dr. Dianne H. Cook and Dr. Jay-Lin Jane.

A special thank of mine goes to Professor Yukihiro Ozaki for his help and guidance at the very beginning of my academic path. His professional knowledge, enthusiasm and resolute dedication to scientific research inspire me to start my own journey of research. I am also very grateful to my colleagues Dr. Lidia Esteve Agelet, Dr. Gretchen Mosher, Mariana Soto, Dr. Benoit Igne and Dr. Igor V. Kovalenko, who listened to my thoughts and offered advice and supports.

Each member in my family has been a great support to me. I want to thank my twin sister Baijing Cao for her support and company through my entire life despite the distance. Your love makes me feel that I am never alone. I want to especially acknowledge my parents for their undivided support and encouragement.

Finally, I want to acknowledge Connie Hardy, Glen Rippke and Howard Shepherd for their contributions that make grain quality lab a big warm family to me.

ABSTRACT

Calibration optimization in near infrared spectroscopy (NIRS) is a complex process, requiring long-term database maintenance and model update by including new variations. A sample selection procedure was introduced to identify the number and choice of samples required in a NIRS calibration model. The example case is the determinations of moisture, protein and oil contents in whole soybeans. The original large database is composed of soybean NIR transmittance spectra (n>8,000) across crop years (2001-2011), varieties and locations. Uniform random, Kennard-Stone and D-optimal algorithms were compared for calibration sample selection. The optimal models based on calibration set selected by uniform random method outperformed the benchmark calibrations using the original dataset with less than 7% of the original dataset for moisture, and less than 30% for protein and oil contents. This procedure was applied to a network of four instruments from two vendors (Foss Infratecs and Bruins OmegAnalyzerGs) to examine the effect of calibration set on calibration transfer. Calibration models of protein and oil contents based on the smallest and optimal number of representative datasets (about 10% and 35% for protein and oil, respectively) were transferred across instrument units of the same brand. Results showed the effectiveness of post-regression slope and bias correction on standardizing predicted values by models built on calibration subsets. Calibrations ($n \approx 120$) built on the selected master instruments were used to evaluate their robustness against temperature fluctuation as an external perturbation. Different temperature compensation approaches were applied to incorporate information of five well-selected perturbed samples. The extended global model and difference augmentation method successfully removed the temperature effect and reduced SEPs on both Bruins (SEPs=0.60% and 0.47% for protein and oil, respectively) and Infratec (SEPs=0.57% and 0.46%, respectively) instruments. Improvements on the predictions of regular samples from crop year 2011 have also been examined with SEPs of 0.51% and 0.34% for protein and oil, respectively on Bruins instrument, and SEPs of 0.52% and 0.34%, respectively on Infratec instrument. Only one or two more PCs were used in the compensated models.

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

In the quality measurement system, the application of near infrared spectroscopy (NIRS) has provided rapid, non-destructive and accurate control in the field of agriculture, pharmaceuticals and food industry. In grain industry, NIRS has a long history in postharvest quality control and real-time quality monitoring during handling and processing (Shenk and Westerhaus 1985; Williams and Norris 1987; Singh, Paliwal et al. 2006). NIRS has been successfully applied in both quantitative analyses as chemical composition determinations and qualitative analyses (discriminant analysis).

NIR has been widely applied in quality control (QC) within grain industry for a very long time. It enables qualitative and quantitative assessment of different types of grains via spectral information and multivariate calibration models. Those models were usually used to determine grain constituents such as moisture, protein and oil. NIR has been approved and used in the Official Inspection System for wheat protein and soybean protein and oil determinations since 1994 (Pierce, Funk et al. 1996). This technique provides fast prediction values of major contents in grain products and reduces time and cost at all stages of production, storage and transport (Osborne 2000). It has been proved that NIR is able to monitor changes occurring in the grain seed during storage according to the spectral difference (Cassells, Reuss et al. 2007). NIR provides a more cost effective and rapid way to measure the quality conditions to adjust the storage management in advance to preserve grain quality and market value.

The application of NIRS is based on a calibration model, which built a mathematical relationship between the absorption spectra and the factor of interest. The process of searching for the best chemometric approaches (linear or non-linear) to interpret the spectra and improve the predictive ability on future samples is called calibration optimization. The motivation of chemometrics is to analyze data to provide chemical knowledge of product contents. Calibration model requires spectra measurements of samples from a population that includes all variances in future prediction. A population is the set of all measurements that covers the characteristics of samples. In agricultural products, the variances could come from the differences of component concentrations, variety, locations, crop years and other external perturbations. Variances from sources rather than the factor of interest could disturb the calibration process and reduce the prediction accuracy. The goal of calibration optimization is to eliminate these effects to a minimum. For agricultural products with complex compositions, it is difficult or sometimes impossible to obtain data points that match with experimental design. On the other hand, a multivariate calibration model is intended to be used for a period of time. However, the changes in sample variations that attribute to different crop years could lead to inaccurate prediction if the new variation is not modeled. The most important things about sample assembly for NIRS work are 1) identify the sources of variance likely to affect the spectra, and 2) assemble samples that contribute this variance, ideally replicated at least 4 times (i.e. 10 different samples that provide the most important sources of variance and fewer samples that offer other forms of variance. The calibration set that is sufficient to develop and evaluate a calibration model should provide all of the likely sources of variance. Quality of grain products is affected by environmental conditions such as climate changes, drought or frost and can vary significantly from year-to-year. Consequently, a procedure of calibration maintenance and updating should be developed. This procedure is to optimize the calibration model with the justification of the number and choice of samples required to calibration an NIRS instrument, and then applied to a multi-model network of NIRS analyzers to meet the standards with one calibration constructed on the master instrument.

In this study, a large population of soybean samples of different varieties is accumulated from crop years 1996-2011 from all over the world and analyzed in Grain Quality Laboratory (Ames, Iowa, USA). This large database provides valuable source of all kinds of variations and ingredient quality factors. However, a large dataset is not computationally effective and would cause problems when using the routine calibration methods for regression. With the inclusion of significantly large number of calibration samples, the risk of involving noise also increases, which might lead to inadequate models (Sáiz-Abajo, Mevik et al. 2005). In order to get robust model and reasonable prediction, a representative subset could be selected to cover the original range of interest components and be balance in the calibration space. By selecting representative, well-distributed data for calibration set from abundant training data, fewer samples are needed in the calibration set without losing the prediction accuracy. The chemical analysis with the traditional methods, which is often more expensive and not operating efficient, can be done only to the selected samples. Then, the life cycle cost of NIRS instrument and calibrations could be reduced.

In practical application, multiple copies of instrument of the same or similar model are distributed in different locations. These instruments form networks in a measurement system and are able to share the same calibration model. In this case, a calibration model is constructed on a selected master instrument and applied to other secondary instruments. The calibration model needs to be adjusted due to the instrumental differences and varying measurement conditions. This process is called calibration transfer or instrument standardization. A network of four instruments from two vendors was formed in this research to examine the effect of different calibration sets on calibration transfer.

Most calibration efforts have been limited to laboratory conditions, while in reality (for example, during the grain handling process), the measurement conditions vary considerably and affect the spectral data with respect to all kinds of external perturbations. In this research, temperature fluctuation was examined as a common environmental factor that possesses variations to NIR spectra. The objective is to develop low-cost and simple temperature compensation methods and build robust calibration models against temperature changes.

1.2 Purpose of Research

The research will develop a procedure based on the correct use of various tools and methods now published for NIR calibration into a workable methodology for routine industrial use, with which users of various skill levels can work. For this purpose, all the environmental and external conditions need to be taken into account, as well as the errors from sampling. With the procedure developed, the financial officer would be able to estimate the life cycle of the cost of applying NIR to a given problem.

The objectives of this research are 1) to determine if the calibration model built on the subset selected from the whole sample set as calibration set has equal or even better prediction ability as using the entire dataset; 2) to identify the optimal sample size and sample selection procedure to construct robust calibration models from calibration subsets; 3) identify the effect of sample size on calibration transfer procedure; 4) enhance the robustness of calibration against external perturbations.

1.3 Organization of Dissertation

This research addressed some important issues in the application of NIR technology in agriculture – robust calibration, accurate prediction and affordable expense. The following chapters will present approaches to optimize calibration models on the prediction of whole soybean grain for moisture, protein and oil contents. The first chapter will introduce and review current chemometric methods on calibration optimization of near infrared models. In the second chapter, sample selection methods will be applied to identify proper numbers and samples in calibration set to construct a robust model. The third chapter applies the sample selection procedure developed in chapter two on the calibration transfer among NIR networks. The fourth chapter deals with the temperature effect on NIR spectra and calibration, which could be corrected by temperature compensation methods. The fifth chapter gives economic analysis on the reduced cost of using less calibration samples with improved prediction performance. The methodologies applied and internal relationships among chapters are shown in Figure 1.

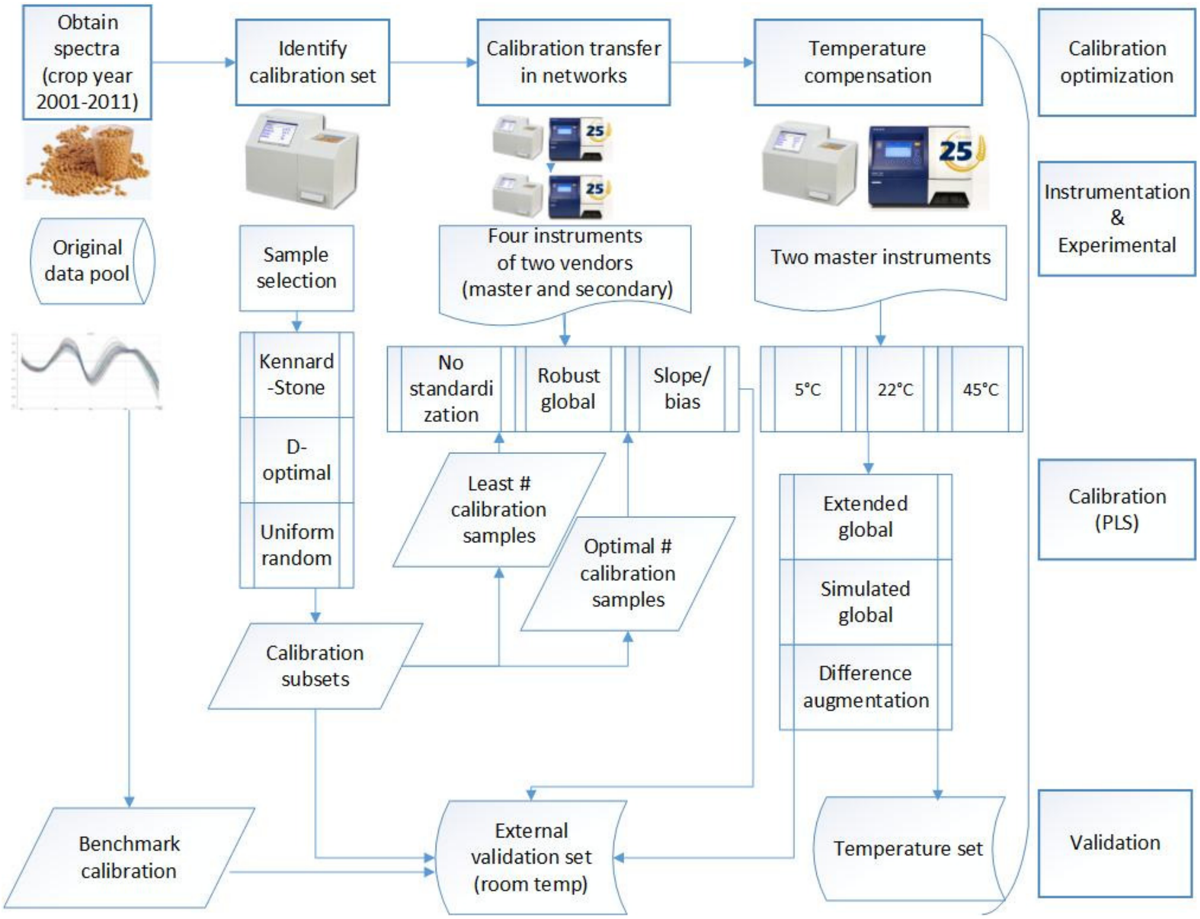


Figure 1. Flow chart depicting the applied methodologies

1.4 Literature Review

1.4.1 Regression methods

Regression methods build mathematic relationship to link the spectrum to quantifiable properties of the samples. The Beer-Lamber's law is the fundamental principle and simplest way of constructing a regression line. Several commonly used regression methods are described briefly in this section. The proper regression approach for multivariate calibration should be selected depending on the data structure on a rational basis.

1.4.1.1 Classical least squares (CLS)

The objective of least squares regression methods is to model and estimate the relationship between a scalar dependent variable **y** and a vector **X** of explanatory variables. In spectroscopy, the classical least squares (CLS) calibration assumes that a measured spectrum is a sum of individual pure constituent spectra weighted by the concentration of the analytes. It requires quantitative knowledge of active individual component in sample, which is only suitable to simple measurement system (Martens and Naes 1992). The CLS model can be written as:

$$\mathbf{A} = \mathbf{CK} + \mathbf{E_A} \qquad (0.1)$$

where **A** is the spectral matrix ($n \times p$) for n samples and p variables, C is the matrix of concentration values of factor of interest, K represents the matrix of pure component spectra at unit concentration and an error term $\mathbf{E_A}$. The least-square solution of K is calculated as:

$$\hat{\mathbf{K}} = \left(\mathbf{C^T C}\right)^{-1} \mathbf{C^T A} = \mathbf{C^+ A} \qquad (0.2)$$

where $\mathbf{C^+}$ is the pseudoinverse of **C** (Haaland and Melgaard 2000).

This method is optimal when the errors are normally distributed. However, CLS is extremely sensitive to outliers that are observations do not yield to the linear pattern formed by the majority of the data. Moreover, the impact of noise could make this method unstable.

1.4.1.2 Multiple linear regression (MLR)

The multiple linear regression (MLR) models assume that in addition to the p independent x-variables, a response variable y is measured, which can be explained as a linear combination of the *x*-variables. The prediction of the factor of interest y_j can be described as:

$$y_j = b_0 + \sum_{i=1}^{k} b_i x_i + e_{i,j} \tag{0.3}$$

where b_i is the computed coefficient, x_i represents the absorbance at each wavelength and $e_{i,j}$ is the error. MLR has been successfully applied to discrete situations in terms of wavelength (Mittermayr, Tan et al. 2001).With proper variable selection method, MLR analysis of a few well-selected variables was able to construct robust equations in low dimensions. The selection is based on the predictive ability of the wavelength.

When MLR is used to construct a predictive model based on spectral data as input and a concentration of factor of interest as output, the method is referred to the inverse least squares (ILS). It uses the inverse form of Beer-Lamber's law. In an inverse calibration, *y* is predicted by fitting the model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{0.4}$$

The regression coefficients can be computed in the form

$$\hat{\mathbf{B}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} \tag{0.5}$$

Then, the concentration of an unknown sample could be calculated from the spectral data $\boldsymbol{x}$

$$\hat{\mathbf{y}} = \mathbf{x}^{\mathbf{T}}\hat{\mathbf{B}} \tag{0.6}$$

1.4.1.3 Principal component regression (PCR)

PCR and PLS models are both based on the spectra of n calibration samples with p variables (wavelengths) from the matrix $\mathbf{X}_{(n\times p)}$ and the vector $\mathbf{c}_{(n\times 1)}$ of the concentration of the factor of interest (Ferre, Brown et al. 2001).

The first step is to perform principal component analysis (PCA) on the spectral data $\mathbf{X}_{(n\times p)}$ (Næs and Martens 1988). Then, the model coefficients $\mathbf{b}_{(p\times 1)}$ for A regression latent variables could be calculated as:

$$\mathbf{b} = \mathbf{R}_{\mathbf{A}}^{+}\mathbf{c} \tag{0.7}$$

The predicted concentration of the factor of interest in a sample whose spectrum is $\mathbf{r}_{(p\times 1)}$ could be presented as:

$$c_A = \mathbf{r}^T\mathbf{b} \tag{0.8}$$

PCA has its advantages on decreasing the dimensions of spectral data and suppressing the spectral colinearity. The problem of PCR is that the principal components describing the spectral data best may not be the optimal PCs for predicting the factor of interest of the unknown samples.

1.4.1.4 Partial least squares regression (PLS)

Partial Least Square (PLS) regression method is the most commonly used regression algorithm in the field of chemometrics in spectroscopy (Burns and Ciurczak 2007). In addition to the first PCA step of determining latent variables, PLS establishes a linear relationship between the spectral data **X** and reference values **y** that maximize their covariance. In PLS, the **X** matrix is transformed to new variables as scores (**T**), loadings (**P**), and vectors called weights (**W**). The scores are orthogonal and estimated as linear combinations of the original variables with the coefficients –weights. Weights are calculated from **X** matrix directly and used to compute T by formula:

$$\mathbf{W}^{*} = \mathbf{W}\left(\mathbf{P}^{'}\mathbf{W}\right)^{-1} \tag{0.9}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W}^{*} \tag{0.10}$$

Then, **X** could be summarized as

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{'} + \mathbf{E} \tag{0.11}$$

where **E** represents the X-residuals. Compared with PCR, PLS takes the reference values in to account to model the information in **X** matrix that describes the factor of interest best. More details about the theoretical explanations of PCR and PLS can be found in reference (Wold, Sjöström et al. 2001). The main advantage of PLS over PCR is to reduce the complexity of the models by using less principal components that contains more related information. This technique was first successfully applied to NIR data in 1982 (Martens and Jensen 1982). In agriculture, it was applied to grain samples by Shenk (Shenk 1991) and to determine milk constituents (Šašic and Ozaki 2001).

1.4.1.5 Artificial neural networks (ANNs)

The term artificial neural networks (ANNs) or simply neural networks (NNs) encompass a family of nonlinear computational methods that were inspired by the functioning of human brain (Marini 2009). NNs have been used as a computational tool for modeling extremely complex functions. It constructs the relationship between a set of inputs and a corresponding set of outputs. Attempted to capture and imitate the biological features of human brain, it could be simply represented by a black-box that receives multiple inputs (**x**) and produces multiple outputs (**y**). This black-box provides a nonlinear relationship between an **x** and a **y** vector. With regards to its functionality, since it is developed to mimic the computational properties of the brain, ANN poses the characteristics such as adaptivity, noise (data), and fault (neurons and connections lost) tolerance (Basheer and Hajmeer 2000). ANN is composed of an input layer, a hidden layer in the case of supposed non-linear relations and one output layer of neurons. The input layer consists of the input variables which usually come from the observable variables such as spectral intensities of the training samples. The output layer would usually be a chemical content value, comprises a layer of artificial neurons. The hidden layer lies in between represents the modeling (training) process allowing output values to be obtained from the input layer of neurons. Within this framework, the data are automatically learnt from an associated set of values by means of chosen training functions. During this training process, a set of parameters known as weights was tuned. In the case study of spectroscopy, the input variables are spectra, the principal component of the spectra or any other forms of data compression, while properties of chemical compounds and multicomponent mixtures for the outputs. A schematic diagram of typical structure of ANNs is shown in Figure 2.

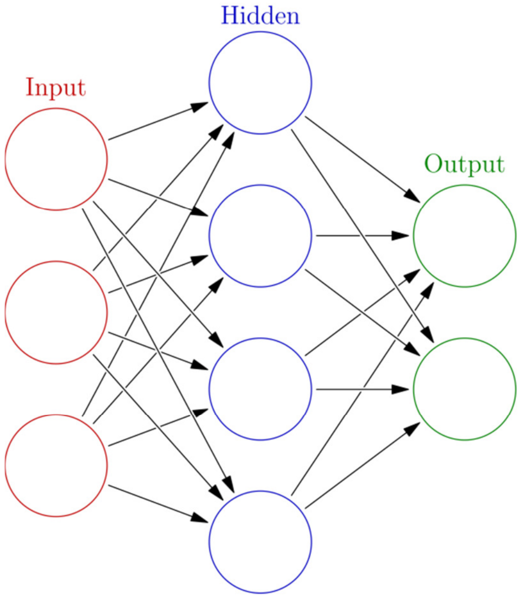


Figure 2. Typical structure of a neural net with three input, one hidden layer and two output layers

Activation function sums the product of the output from each unit and the weight with which it is connected to the current unit (Swingler 1996).

$$o_j = f\left(\sum_{i=1}^{m} w_{ji} o_i\right) \qquad (0.12)$$

where m denotes the number of units in the current layer, o_j denotes the output from unit *j*, w_{ji} denotes the strength of the weight from unit *i* to unit *j* on the next layer. $f(\bullet)$ is the network activation function (or called transfer function). The most used transfer function is the logistic function:

$$f(x) = \left[1 + e^{\left(-\frac{x}{\theta}\right)}\right]^{-1} \qquad (0.13)$$

where x is the weighted summation of input signals, θ refers to the gain (Burns and Ciurczak 2007). In this function, *x* is squashed into the range from zero to one and the derivate makes very small changes at either end of the range and larger changes in the middle. Both the logistic and the tanh functions are commonly used in NIRS, which could be called as sigmoidal functions. However, sine and simple linear functions are also widely used. In NIRS analysis, PCA or PLS scores are always used as input variables in ANN models. The optimal number of layers for network and that of neurons for each layer vary in the case studies. By using the PC scores as input, the robustness of the ANN calibrations could be improved and computation is reduced. Study (Janik, Cozzolino et al. 2007) also proved that prediction using ANN with PLS scores as input improved the capability of modeling non-linear relationships. Nevertheless, the number of input nodes needs to be optimized if the PCA scores are used as input in order to avoid including excessive irrelevant scores and redundant data, which leads to over-fitting. This value is typically higher than the number of PCs used in linear models. Inputs are transferred to neurons in the next layer with their respective randomly assigned weights. The net summation of all weighted signals is calculated:

$$Net_j = \sum_i w_{ij} x_i + \theta_j \qquad (0.14)$$

x_i denotes the different inputs for a neuron *j*, w_{ij} refers the weight of the connection through which signal x_i enters neuron *j*. Then, Net_j is related to the neuron output values through a transfer function. In a three layer network, outputs of the hidden layer are multiplied by a weight and sent to the output layer where they are summed and applied to the transfer function to obtain predictions. These predictions are compared with reference values and a calibration error is determined. If the value is greater than a threshold predetermined by the user, an optimization process such as back-propagation (BP) learning rule (Sun 2009) could be applied to adapt the weights and reduce the calibration error. The scheme of the BP algorithm is update the weights iteratively until the error criterion reaches a minimum.

During the ANN training iteration, a set of parameters need to be defined, such as the weighted randomization range and seed, regulate the speed and the stability of error convergence. The choice of the initial values of the connection weights has an important impact on the convergence properties of the network (Marini 2009). It was suggested to initialize the weights using a uniform distribution that has zero mean and a variance that is equal to the reciprocal of the number of connection to the units. Another vital problem is to set the stopping criterion. The training error is supposed to decrease with the increasing number of training epochs and reaches an asymptotic value. However, the network might learn a relevant part of noise together with the systematic information during the iteration. Then, the corresponding model is said to lack of generalizing ability if an unknown sample is presented to the network. Thereby, a validation procedure could be applied by using an external set to obtain the optimal number of iterations. The training will stop when the lowest value of the generalization error is reached.

With regarding to select the best architecture, for practical purposes, it may not be possible to derive the optimal number of hidden layers only based on the theory. In practice, the risk of over-fitting is a crucial issue that should be eliminated. This could also be accomplished by test a series of numbers of hidden layer on an external validation set.

With respect to target error (Swingler 1996), the number of samples in the training set (n) could be approximated by:

$$n \geq \left(\frac{w}{\varepsilon} \right) \tag{0.15}$$

if $0 < \varepsilon \leq 1/8$, where w denotes the number of weights, ε the low error limit. With large enough data set, then

$$w \approx \varepsilon n \tag{0.16}$$

The number of samples needed when there are sufficient samples would be discussed in more depth in the following section.

1.4.2 Calibration set selection

There is no rule of thumb to determine the optimum number of selected samples. This number depends on the complexity of the corrections (wavelength shift, intensity offset) and on the algorithms used. Research showed that a sample set has a ratio of samples to variables larger than four was considered as large (Naes, Irgens et al. 1986). Different subset sizes should be selected and assessed for its effect on calibration robustness.

Sample selection involves identification of all sources of variance likely to be encountered in future analysis including sample source (growing location and season) and the range of composition in constituents or parameters to be tested. Selection of sample with normal distribution will cause the results of subsequent analysis to regress toward the mean. Sample selection usually depends on sample information, variable selection, X-Y relation outliers, Y-representative and sample residual in the model. Sample information includes the refractive index of liquids changes with concentration and inter-correlation between components (moisture content/particle size with components of interest). The distribution of sample component concentration (y values) should be as evenly as possible.

Adding/eliminating more wavelengths (variables) to the spectrum is analogous to adding/eliminating samples to the training set. As the samples in training set, if the wavelengths are exact duplicates or there are large numbers of wavelengths which are similar, wavelengths do not expand the dimension of S, nor help S span V (Hildrum 1992).

1.4.2.1 Sample Selection Algorithms

A handful samples will probably not carry enough weight. The fundamental principle of sampling states that each sample is representative of the total material (population) from which it is taken (Haswell 1992). All units present in the population should have an equal probability of existing in the representative sample. Several sample selection methods are available in literature.

1) Uniform random selection

Random selection is the simplest way to choose samples from a large population. However, this would lead to another normal distribution of reference values in calibration set. This distribution has more samples around the means, and less extreme values. For better prediction, the distribution of the calibration samples should be as uniform as possible. In this case, we use the method of uniform random selection. The total range of reference values are divided into several intervals. In each interval, we randomly select certain amount of samples. Then combine them together, we should get a relatively evenly distributed calibration set.

2) Selecting the samples with the highest orthogonal leverage values

This function of this algorithm (Wang, Veltkamp et al. 1991) was designed to select subsets of spectra for use in instrument standardization transform development based on sample multivariate leverage. Leverage is an observation concept which concerns the position of the observation's "independent variables" relative to the others. High leverage samples have strong effect on x and y loadings. Thus outliers must be detected prior to the subset selection; else they would be selected in the high leverage subset. The information contained in the selected samples is then removed from the rest of the samples by a linear transformation so that they are all orthogonal to this selected sample. This procedure stops when the desired number of samples has obtained. The algorithm steps are as follow:

i. Calculate the hat matrix **H** for calibration set, of which the diagonal elements are the leverages that describes the influence each observation has on the fitting value of that same observation. For linear models, the hat matrix

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad (0.17)$$

ii. Select the sample with the highest leverage value (maximum h_{ii}), and orthogonalize the row selected against every other row of sample spectra to obtain a new X, i.e., calculate the linear transformation

$$\mathbf{x}_j' = \alpha \mathbf{x}_i + \beta \mathbf{x}_j \text{ for } j \neq i \qquad (0.18)$$

subject to

$$\mathbf{x}_j' \bullet \mathbf{x}_i = 0$$

iii. Repeat the previous two steps until the number of selected samples reach the number desired.

3) Shenk-Westerhaus method

This is also known as the patent method (Shenk 1991; Shenk 1991). The selecting function of WinISI software (Infrasoft International, State College, PA) was designed to choose samples that maximize the global and neighborhood standardized distances (global and neighborhood H) on a PCA using Mahalanobis distance. The H limits are parameters that must be carefully set. For example, in a forage and grain analysis, a standardized H value of

3.0 was used to exclude outliers and a minimum H value of 0.6 determined the neighbor samples. The final population is even and symmetric, with fewer samples at the edge.

4) Kennard-Stone method

The Kennard-Stone algorithm (Kennard and Stone 1969) is designed to select samples sequentially which are uniformly distributed over the object space by choosing the samples that maximize the Euclidean distances between each other (designed for surface response experimental plans). The distances are computed between the sample characteristics (e.g., y-values or spectra). The first two farthest samples are selected. Then, the third sample is selected as the one farthest from the first two samples. Keep choosing samples one by one from the remaining subset, which is the farthest from all the previously selected till reach the designed number of training set. Supposing that k objects have already been selected (k<n, n is the number of samples), the (k+1)th object in the calibration set is chosen using the criterion

$$\max_{k<r\leq n}\left(\min\left(d^2{}_{1r,}d^2{}_{2r,\ldots}d^2{}_{kr}\right)\right) \tag{0.19}$$

where

$$d_{jr}=\left\|x_j-x_r\right\|=\sqrt{\sum_{j=1}^{k}\left(x_j-x_r\right)^2} \tag{0.20}$$

It denotes the Eucidean distance from a candidate object r, not yet in the representative set, to the jth already selected object. This algorithm chooses the sample that presents the largest minimum distance with respect to any sample already selected at each successive iteration.

An improved method based on Kennard-Stone algorithm was proposed for dividing a data set into calibration and validation subsets(Galvao, Araujo et al. 2005). This method defined the Eucidean distances of both spectral data and reference data. By dividing dx and by their maximum values in the data set, equal importance was assigned to x and y spaces. A normalized xy distance d_{xy} was used instead of d_x alone in KS algorithm.

5) Duplex method

The Duplex algorithm (Snee 1977) was first designed to split the data into calibration and prediction sets, which cover approximately the same region and have similar distribution. It is a modification of KS algorithm to achieve randomization in sample selection. Similar to KS algorithm, the two points that are furthest away from each other are selected in the calibration set at first. Then, the sample that is furthest away in the remaining dataset is assigned to the validation set. In the third step, point furthest away from the selected calibration samples is included in the calibration set. This procedure repeats till representative samples in calibration and validation set reach an equal size.

6) Cluster analysis

The first step of this method is to perform a PCA on all the spectral data in the entire data set and choose the number of principal components. Then, divide samples into several clusters based on the similarity (eg. Euclidian distance) of their PCA score values. The number of clusters is set according to the expected number of samples in the calibration set. Select one sample from each cluster using complete linkage (farthest points define the inter-cluster distance) (Næs 1987; Isaksson and Næs 1990).

7) Puchwein method

Sort samples according to Mahalanobis distance to the center of the set. Select the extreme point. Exclude points that are closer to the selected point by a limiting distance. The sample that is most extreme among the remaining points is selected. Then repeat the procedure until there is no data point left. The limiting distance is the factor that controls the number of samples being selected (Puchwein 1988).

8) D-optimal method

D-optimal designs (de Aguiar, Bourguignon et al. 1995; Ferre and Rius 1997) are used for experimental design when the experimental region does not have a regular shape. Its principle is to maximize the determinant of the variance-covariance matrix $|X'X|$, where X is the training set matrix (p selected samples, m wavelengths). This determinant is maximal when the selected samples span the space of the whole data set. This algorithm starts with a large data set and chooses samples iteratively that create a maximum increase in the variance-covariance matrix $|X'X|$, and stops when choosing a sample no longer increase the variance-covariance matrix $|X'X|$. We run this algorithm several times to avoid getting a local optimal subset.

In this research, the procedure of selecting calibration set is depicted in Figure 3. For comparison purpose, calibration based on the whole dataset was used as benchmark method. Calibration subsets with increasing numbers of samples were used to build models for an external validation set from the latest crop year.

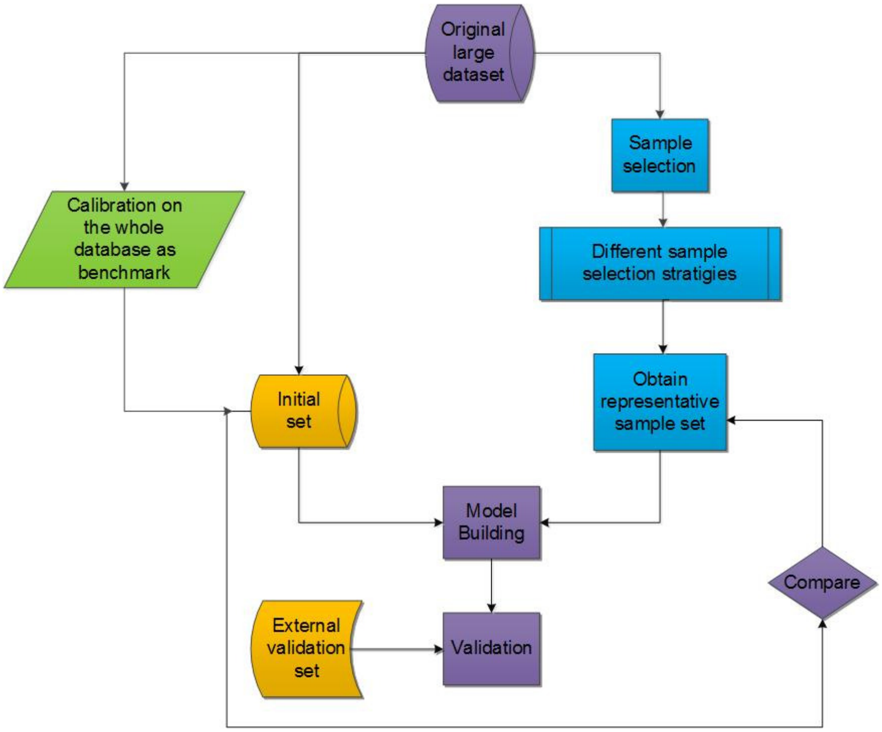


Figure 3. Flowchart of calibration sample selection procedure

1.4.3 Variable selection

As another important aspect of calibration optimization, variable selection has the same goal as sample selection with respect to reduce the complexity of the calibration model. A great deal of studies have shown that the construction of calibration model on selected relevant predictors from the overall variables improves the robustness of calibration against to uninformative wavelength regions (Westad and Marten 2000; Höskuldsson 2001; Abrahamsson, Johansson et al. 2003). The inclusion of all variables could affect the prediction properties and lead to suboptimal models in the presence of highly correlated variables. It is necessary to select appropriate variable to develop a calibration model that gives an adequate and representative description for use in prediction (Gemperline 2006). To evaluate the variable subsets on the regression results, a validation set is usually used based on minimizing the prediction-error.

Various variable selection methods have been proposed to identify informative wavelength regions such as interval PLS (iPLS) (Norgaard, Saudland et al. 2000), genetic algorithm (GA) (Lucasius, Beckers et al. 1994; Smith and Gemperline 2000), uninformative variable elimination (UVE-PLS) (KOSHOUBU, IWATA et al. 2001; Cai, Li et al. 2008) and Monte Carlo uninformative variable elimination (MC-UVE) (Li, Liang et al. 2009). iPLS searches for the combination of relevant information in different spectral subdivisions. GA is based on the principle of genetics and natural selection. It also provides a way for data compression to select input variables for ANN (Despagne and Massart 1998; Chalus, Walter et al. 2007).

Variable selection will not be the main focus in this research compared to calibration sample selection. Due to the characteristics of grain products and condensed wavelength region (shortwave) of transmittance instruments utilized in this paper, eliminating the number of samples takes the first priority in optimization.

1.4.4 Outlier identification

Outlier detection techniques can be used to predict the “uniqueness” of a sample using H statistic, also known as the Mahalanobis distance. Residuals plots are important in multivariate regression. Leverage corrections of the residuals are plotted as “influence”. Samples with high residual and high leverage are considered as outliers. Leverage and residuals are the most important values for detecting outliers.

In this study, PLS is first carried out (leave-one-out cross validation) on spectra collected on individual instrument within each crop year in order to clean outliers and visual inspection. For those years having few samples, samples were joined to the ones from the next year. The basic tools for outlier detection in these models are based on the influence plot in Unscrambler 9.8 (Camo Software, Woodbridge, NJ) with leverage as x-axis and y-residual as y-axis.

The criteria for deleting outliers in this study were 1) samples with residuals higher than 2; 2) samples with large leverage: Samples with leverage higher than 3 times the average leverage, where the average leverage is calculated as:

H=1/n + number of principal components/n (n is the number of samples) (Faber 1999; Faber 1999)

1.4.5 Calibration model selection

After acquiring a series of calibration models based on different sample sizes (increasing numbers of samples), the selection of the optimal calibration model becomes a critical step in optimization procedure. In a multivariate calibration, sample size has a substantial impact in achieving statistical significance, both in small and large sample sizes (Anderson, Hair et al. 2006). For calibration set with less samples included, the complexity of the multivariate technique may easily result in either (1) too little statistical power for the test to realistically identify significant results or (2) too easily “over-fitting” the data such that the results are artificially good because they fit the sample yet provide no generalizability. On the other hand, for large sample size, the statistical test becomes sensitive in terms of the significance of results.

It has been demonstrated the use of experimental design could help choose the optimal calibration model (Flaten and Walmsley 2003; Flåten and Walmsley 2004). The experimental design approach was used to choose all parameter setting including pretreatment methods (categorical variables), the number of components, calibration subset and variable subset selection.

To compare estimates of prediction error of the optimal calibration set and the whole sample set, one single validation set of N_p samples with known **x** and y is used to predict y from **x** using each of the calibration models. Since the true y is known, this gives a set of N_p prediction errors ($\hat{y} - y$) for each method.

$$\left(\hat{y}_{ij} - y_{ij}\right)^2 = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ij} \tag{0.21}$$

Simply without interaction,

$$\left(\hat{y}_{ij} - y_{ij}\right)^2 = \mu + \alpha_i + \beta_j + e_{ij} \tag{0.22}$$

Multiple comparisons are made among different calibration models with respect to prediction of external spectral samples. For unknown samples, the prediction error can be determined as root mean square error of prediction (RMSEP):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}} \tag{0.23}$$

where n is the number of unknown samples in the validation. While RMSEP measures the accuracy of a prediction, the square error of prediction (SEP) measures precision and can be calculated as:

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}-y\right)^2 - \frac{\sum_{i=1}^{n}\left(\hat{y}-y\right)^2}{n}}{n-1}} \tag{0.24}$$

where Bias is calculated as

$$Bias = \frac{\sum_{i=1}^{n}\left(\hat{y}-y\right)}{n} \tag{0.25}$$

Then, the relationship among SEP, Bias and RMSEP is depicted as:

$$RMSEP^2 \approx SEP^2 + Bias^2 \tag{0.26}$$

The predictive ability of the calibration model often reaches a minimum RMSEP at the optimum number of factors and begins to increase thereafter. As shown in Figure 4, when evaluating the bias of the model with respect to prediction error, there is a trade-off of variance for prediction estimates with respect to bias. This means, with the increasing complexity of calibration model, the bias decreases at a sacrifice of a variance increase.

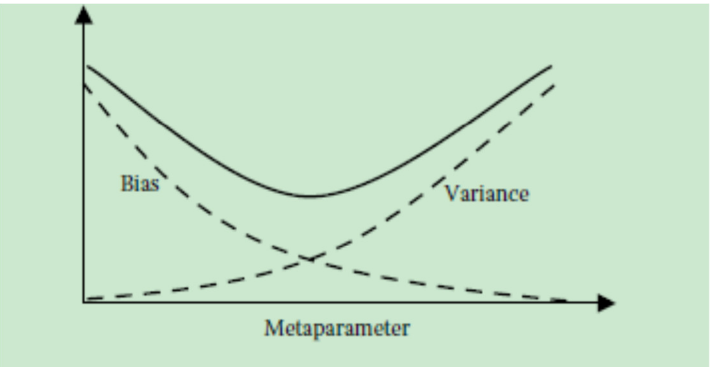


Figure 4. A generic situation for model determination showing the bias/variance trade-off with selection of metaparameter (Gemperline 2006)

Tukey's method (Indahl and Naes 1998) for multiple comparisons can be used for this purpose, which is a conservative test that compares all differences between pairs of methods with the same critical rules. There are two assumptions of Tukey's test needs to be met: 1) the observations being tested are independent; 2) there is equal variation across observations (homoscedasticity). The difference can be expressed as the absolute residual:

$$d\left(y,\hat{y}\right)=\left|y-\hat{y}\right| \qquad (0.27)$$

Multiple-way fixed effects ANOVA without interactions can be used to test differences among the performances of calibration models. It is also necessary to exam whether the residuals of the ANOVA are normally distributed.

For comparison of two SEPs, the appropriate methodology to test their differences is as below

$$\kappa=1+\frac{2(1-r^2)t^2_{(N_P-2),0.025}}{N_P-2} \qquad (0.28)$$

where r is the correlation coefficient between the two sets of prediction errors, $t_{(N_P-2),0.025}$ is the upper 2.5% percentile of a t-distribution with N_p-2 degrees of freedom. Then, find

$$L=\sqrt{\left[\kappa+\sqrt{(\kappa^2-1)}\right]} \qquad (0.29)$$

Now

$$\frac{SEP_1}{SEP_2}\times\frac{1}{L}\ and\ \frac{SEP_1}{SEP_2}\times L \qquad (0.30)$$

Give the lower and upper limits of a 95% confidence interval for the ratio of the true standard deviations. If the interval includes 1, the standard deviations (SEPs) are not significantly different at the 5% level.

1.4.6 Calibration transfer and instrument standardization

The aim of instrument standardization is to build robust models of calibration and prediction. Calibration transfer enables robust models to be employed on a class of instruments in filed or in-line. The problems are the existence of instrument-to-instrument variations and their small differences in wavelength resolution and detector sensitivity. Moreover, it gets more complicated when time-dependent instrumental drift occurs. Due to all these reasons, instrument standardization is needed in application.

Multivariate calibration has become routine for extracting chemical information from spectroscopic signals (Brown 2009). The most commonly used multivariate methods for chemical analysis are partial least squares (PLS) regression (Zhang, Small et al. 2002) and principal component regression (PCR). A robust calibration model is supposed to be used for extended periods of time. In this case, samples to be predicted in the future would be measured under a different environmental condition from samples in the calibration set. However, with a new condition of spectra collection, the changes of variations in the spectra could lead to invalid prediction results. There are three situations that could induce model inconsequent (Feudale, Woody et al. 2002) : 1) changes in the physical and/or chemical constitution of the samples; 2) changes in the instrumental response function; 3) changes in the environment factor over time. For agricultural products, these changes would occur when there are fluctuations of temperature or humidity during the storage and spectra collection.

Previous studies have developed various standardization and preprocessing methods for calibration transfer between different systems. While this paper focused on the transfer of near infrared data, standardization methods could also be applied to UV-visible spectrophotometry, fluorescence spectroscopy, Raman spectroscopy and electrochemistry. Usually the instrument where the calibration models was originally made is defined to be the master instrument, while the calibration is to be transferred to as the secondary instrument (Næs, Isaksson et al. 2002).

1.4.6.1 Global models

To build a global model, spectra scanned from more than one instrument are included in the calibration set. By incorporate samples covered a wide range of experimental conditions, expected variations are implemented in the calibration model. The only problem is the source of variation in the future prediction samples need to be strictly controlled. Study (Igne and Hurburgh 2008) showed models built on samples scanned in two instruments of the same brand and two instruments of different brands gave equal or better results than when each instrument was calibrated on its own calibration set.

1.4.6.2 Model updating

To build a new model that contains the new variations could be time-consuming and may be expensive. After model development, calibration needs to be maintained and updated at some point of time. An alternative way is to only add more samples selected from the new incoming set. This would also involve the sample selection process as to add representative samples account for the new source of variances as well as possible. With the additional samples in the calibration set, the model could be more robust to the new measurement conditions and lead to better prediction results. Due to the small number of the new samples, their weight might be too low compared to the existing data set. A possibility would be to give the additional samples more weight to increase their contribution to the model. However, research (Capron, Walczak et al. 2005) showed that the number of new samples included had a larger impact on the performance of the updated models. Moreover, sample selection strategies are preferred to select incoming samples different than existing data set.

Another approach was developed by combining PCA and PLS to examine the similarity of a new unknown sample to the samples had been already defined in the calibration set by Setarehdan et al. (Setarehdan, Soraghan et al. 2002). The entire procedure was depicted in Figure 5.

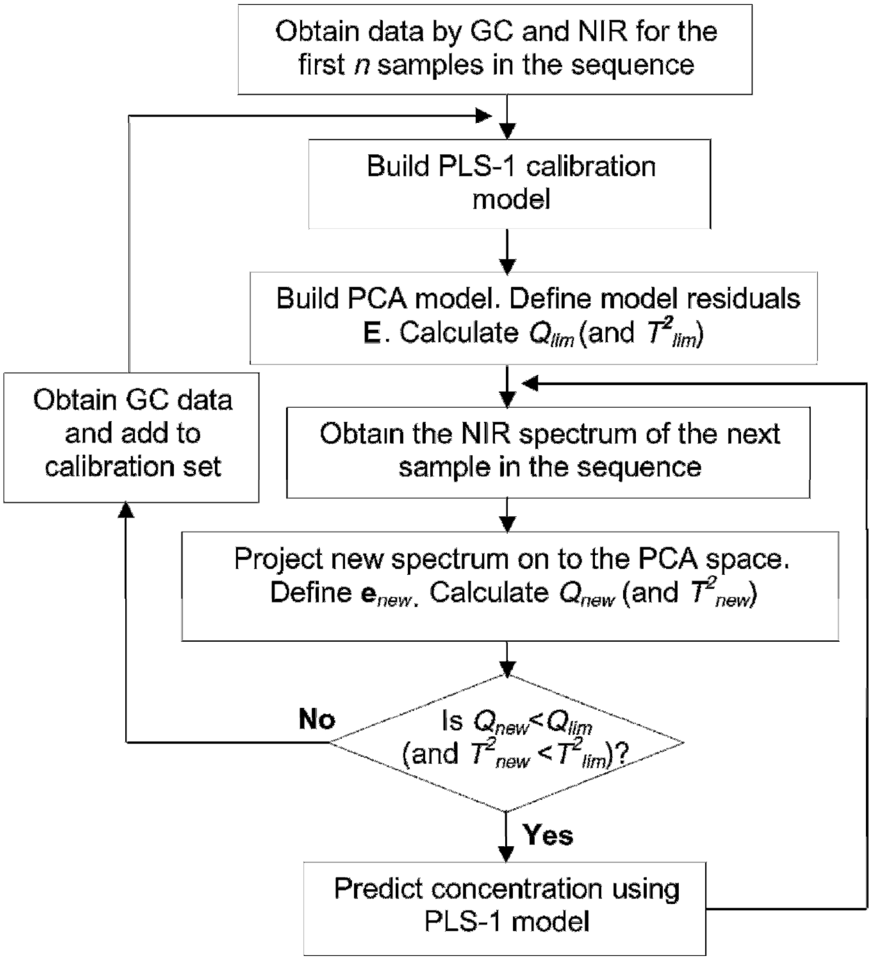


Figure 5. The overall block diagram of the calibration update algorithm (Setarehdan, Soraghan et al. 2002)

This strategy requires the reference values of the first n samples in the calibration set. Q and T^2 statistics were used to determine whether a new input spectrum should be added into the calibration set. Q is the sum of squares of each row of the residual matrix. It describes the amount of the variation in each sample that is not captured by the principal components retained in the model and can be calculated as:

$$\mathbf{Q_i = e_i e_i^T} \tag{0.31}$$

The sum of normalized squared scores, known as Hotelling's T^2 statistics is a measure of the variation of the samples "within" the PCA model, which can be defined as:

$$\mathbf{T_i^2 = t_i \lambda^{-1} t_i^T} \tag{0.32}$$

where $\mathbf{t_i}$ refers to the matrix of score vectors from PCA model, and λ^{-1} is the diagonal matrix containing the inverse of the eigenvalues associated with principal components retained in the model. The results of this study demonstrated an invariant prediction ability of the calibration model built on the initial number of selected samples. This provided the feasibility of building robust calibration on a small, representative calibration set and reduction on the cost and energy of calibration efforts.

1.4.6.3 Instrument standardization methods

Great deals of studies have attempted to propose strategies to deal with variations in the instruments, sometimes even between instruments with different optical configurations (Puigdomènech, Tauler et al. 1997; Wang, Su et al. 2002). Most of the methods standardize the instrument by mathematically manipulating the regression coefficients in the calibration model, the spectral responses or the prediction values. This section states several common standardization methods that are widely applied.

1) Shenk and Westerhaus standardization

This approach was first proposed by Shenk *et al.* and patented (Shenk, Westerhaus et al. 1985; Shenk and Westerhaus 1989). This is the simplest way to correct the changes in spectral responses between the master and secondary instruments directly. This method was originally designed for the standardization of similar monochromator NIR reflectance instruments with a wavelength alignment. The advanced version of this method allows transfer between NIR reflectance instruments with equal resolution. With the measurement of standardization set on both instruments, it consists of two separate steps: 1) wavelength index correction; 2) spectral intensity correction. In the first step, the wavelength scale is corrected by correlating the measurements at wavelength *i* on the master instrument with those located in a small window around *i* on the secondary instrument. For each wavelength *i* in master instrument, a spectral window of neighboring wavelengths on the secondary instrument is chosen to calculate the relationships. A second-order polynomial is fitted to the correlation coefficients, yielding a continuous function across the channels in the processing window. The wavelength corresponding to the maximum of the quadratic function is then selected as the corresponding wavelength of the master instrument. In the second step, the intensity at each wavelength *i* is correlated by simple linear regression:

$$\mathbf{x}_{\mathbf{i,1}} = a_i + b_i \mathbf{x}_{\mathbf{j,2}} \tag{0.33}$$

Where a_i and b_i are the offset and slope, respectively, estimated from the standardization samples measured on both instruments.

Only linear intensity differences can be corrected by this method. The disadvantage of this method is that complex interactions between wavelength shifts and intensity changes cannot be transferred since it assumes that no relationship exists between neighboring correction models. Thus, it only applies to similar instruments and the drift correction within one instrument. Another requirement of this method is a standardization sample set measured on both instruments. Research (Bouveresse, Massart et al. 1994) was conducted to examine the effect of different standardization samples of agricultural products by using this method. Three different kinds of standardization sample sets were tested including samples similar to the agricultural samples, generic standards and pure organic and inorganic chemicals. It stated that difference between spectra obtained on two different instruments depend probably on the optical density range. The standardization samples used to correct differences between the two instruments should be similar to the samples of the prediction set. This research also provided some possible strategies to select standardization samples.

2) Slope and bias correction

Another widely used method for correcting predicted values is the simple univariate slope and bias correction (SBC). In this method, it assumes a linear relationship between the prediction values for spectra measured on the secondary instrument and the prediction values that the obtained by calibration built on the master instrument. The properties of the standardization samples (y-values) were firstly predicted by the original model developed on the master instrument.

$$y_{s,1} = \mathbf{X}_1 b \tag{0.34}$$

$$y_{s,2} = \mathbf{X}_2 b \tag{0.35}$$

A linear regression equation is obtained by plotting spectra collected in the calibration step against those collected in the prediction step using either ordinary least squares or orthogonal least squares. Then the predicted values for the new samples are corrected for the bias (intercept) and slope of the regression line:

$$y_{2,corrected} = bias + slope\,\mathrm{y}_2 \qquad (0.36)$$

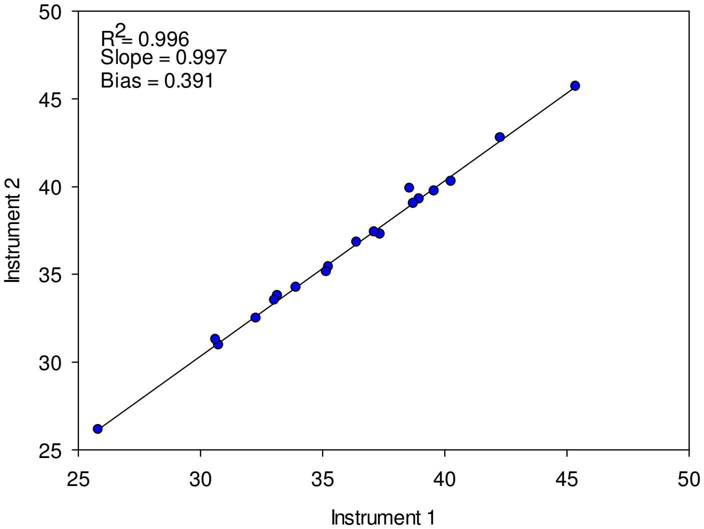


Figure 6. Linear relationship of the prediction values on the standardization samples between two similar instruments

Figure 6 shows an example of the prediction values of the standardization samples collected on two similar instruments. Acquiring the slope and bias values, the prediction on the secondary instrument could be corrected by using calibration model built on the master instrument.

This standardization approach is most often applied between instruments having the same dispersion device. When more complex instrumental differences exist, other standardization methods need to be used. A procedure was proposed to diagnose whether the simple slop/bias correction can be successfully applied based on a statistical F-test (Bouveresse, Hartmann et al. 1996).

Osborne and Fearn (Osborne and Fearn 1983; Osborne and Fearn 1983) utilized this method to transfer calibration equations built for the determination of proteins and moisture in wheat flour. Nine instruments of the same model located in different laboratories were used to collect one sample set and a second sample set in two month later to adjust the bias for a specific component. The results indicated reasonable accuracy of NIR as well as precision. Another collaborative research conducted by Delwiche, Pierce et al. (Delwiche, Pierce et al. 1998) assessed accuracy, repeatability and reproducibility of NIR method for determine crude protein content in whole grain products. In this research, four types of commercialized NIR instruments with various combinations of wavelength region, mode of energy capture, dispersion and treatment of spectral data were used. Twenty two standardization samples and a test set of twelve unknown samples were used for all collaborators. For standardization, bias correction, slope and intercept correction and recalibration with inclusion of standardization sample spectra were examined. The results demonstrated for within-laboratory and between-laboratory variations of the NIR method were equivalent to values reported for the combustion method (AOAC 1995) for wheat. Furthermore, Osborne, Kotwal et al. reported a method using the single sample to correct for bias (Osborne, Kotwal et al. 1999) on reflectance monochromator instruments. This was considered as a simplified procedure of Shenk-Westhaus method (Shenk, Westerhaus et al. 1985) for whole grain samples. In comparison to the set of thirty samples, using single ground wheat confirmed the standardization worked equivalently. The experiment described using one single sample to correct sufficiently for the purpose of combining spectral data sets from different instruments. The only consideration of the single sample standardization is the accuracy of the wavelength axis without the correction by the standardization procedures. The results indicated that the wavelength accuracy of the instruments employed was not a significant factor during standardization.

3) Direct standardization of instrumental response

To directly relate the response of a sample measured with one instrument to its response obtained on another instrument, Wang et al. (Wang, Veltkamp et al. 1991) proposed direct standardization by means of a transformation matrix. The linear relationship is stated by the transformation matrix $\mathbf{F}$ according to:

$$\mathbf{X}_1 = \mathbf{X}_2\mathbf{F} + \mathbf{E} \qquad (0.37)$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are the response matrices of the standardization samples obtained from the master and secondary instruments, respectively. $\mathbf{E}$ represents the residual matrix. The transformation matrix $\mathbf{F}$ is a square matrix and determined by multiplying the generalized inverse of the standardization set obtained in the prediction step by the standardization set obtained in the calibration step:

$$\mathbf{F} = \mathbf{X}_2^{+}\mathbf{X}_1 \qquad (0.38)$$

where $\mathbf{X}_2^{+}$ is the generalized inverse of $\mathbf{X}_2$. Once $\mathbf{F}$ is calculated, the response vector of a new sample x is projected to the original measurement space so that its property values can be predicted with the original model:

$$\hat{\mathbf{x}}^{\mathbf{T}} = \mathbf{x}^{\mathbf{T}}\mathbf{F} \qquad (0.39)$$

The computation of $\mathbf{F}$ assumes that change in response values is caused by the change in the instrument function. However, any variation in the chemical composition of the samples will also be incorporated into the model. Another problem is the number of standard samples has to be at least as large as the rank of $\mathbf{X}_{\mathbf{A}}$, in order to represent all relevant dimensions. Otherwise, it is usually underdetermined with respect to estimating $\mathbf{F}$. When compared with other standardization schemes, DS showed large SEP when the number of subset samples selected as standardization samples is smaller than the rank of $\mathbf{X}_{\mathbf{A}}$ (Wang, Veltkamp et al. 1991). It does not require the master and secondary instruments have the same number of wavelength points. When the number of variables is much larger than the number of standardization samples, there is a high risk of overfitting of the matrix $\mathbf{F}$. Therefore, $\mathbf{F}$ is typically estimated by means of PCR and PLS regression to obtain a least squares solution. Another alternative approach to overcome this problem is to reduce the number of channels involved in the regression, which gives rise to piecewise direct standardization.

4) Piecewise direct standardization

Piecewise direct standardization (PDS) (Wang, Veltkamp et al. 1991) is one of the most widely used transfer methods. In DS, each wavelength of the master spectra is related to all wavelengths of the secondary spectra. When one instrument is shifted along the x-axis with respect to the other instruments, the spectral correlations are usually limited to specific and smaller regions. Thus, PDS is a local alternative approach. In PDS, the response $\mathbf{r}$ of the standardization samples measured at wavelength j on the master instrument is related to the wavelengths located in a small window around j measured on the secondary instrument:

$$\mathbf{r}_{\mathrm{j}} = \mathbf{R}_{\mathrm{j}}\mathbf{b}_{\mathrm{j}} \tag{0.40}$$

where $\mathbf{R}_{\mathrm{j}}$ is the localized response matrix of the transfer samples and $\mathbf{b}_{\mathrm{j}}$ is the vector of transformation coefficients for the *j*th wavelength. The regression vectors calculated for each window in the data are then assembled to form a banded diagonal matrix $\mathbf{F}$ according to:

$$\mathbf{F} = diag\left(\mathbf{b}_1^{\mathrm{T}}, \mathbf{b}_2^{\mathrm{T}}, \ldots, \mathbf{b}_{\mathrm{j}}^{\mathrm{T}}, \ldots, \mathbf{b}_{\mathrm{k}}^{\mathrm{T}}\right) \tag{0.41}$$

where k is the number of spectral channel (wavelength). For any unknown sample, the spectral response could be standardized using the equation (0.42) as described in DS method. When there is not sufficient data to form a complete window, edge effects would occur due to the operation of a moving window. The ends of the spectra are either removed or estimated by extrapolation under these circumstances. The two major advantages of PDS are the use of a small amount of samples in a secondary instrument due to the smaller local rank of moving windows than the whole matrix and its multivariate nature enabling a noise-filtering effect. Non-linearities are better explained by several local multivariate models than a single global local multivariate model.

Wang et al. (Wang, Veltkamp et al. 1991) compared five standardization methods included standardization with the classical calibration model, standardization with the inverse calibration model, DS, PDS and the patented method by Shenk and Westerhaus (Shenk and Westerhaus 1989). 100 mixtures were simulated by a full experiment design (10 by 10) with the first 10 subset samples selected as standardization samples. PDS gave the best results among these different procedures and obtained the smallest SEP with only three subset samples. Results similar to subset recalibration were obtained when the number of subset increased to 6, which was sufficient to obtain a correction for the nonlinear response change.

The effect of window size associated with the number of subset samples on standardization was examined as well. The results indicated that over-fitting may occur with the inclusion of too many non-relevant channels, which was the major problem of direct standardization where the full spectral region is included. In their following research, Wang et al. (Wang, Lysaght et al. 1992) were able to improve the calibration through standardization if an instrument with higher quality in terms of signal-to-noise ratio is standardized. The results demonstrated better SEPs than that of the whole set recalibration due to the utilization of a better calibration through standardization.

PDS also has been applied to situations of standardization with nonlinear external influence as temperature (Wang and Kowalski 1993). A modified version called Continuous Piecewise Direct Standardization (CPDS) (Wulfert, Kok et al. 2000) was proposed to correct the continuous temperature effect. In this method, a CPDS model is built to standardize spectra between calibration sample temperature and various discrete temperatures. Then a polynomial regression is fitted for the values at each position against the temperature difference. Estimation of transformation matrices were obtained for all temperature differences lie in the standardization range.

PDS is commonly used as a reference method to be compared with other novel standardization techniques (Swierenga, Haanstra et al. 1998; Zhang, Small et al. 2002; Honorato, Galvão et al. 2005; Watari and Ozaki 2006; Guenard, Wehlburg et al. 2007; Sohn, Barton et al. 2007; Igne and Hurburgh 2008; Igne, Roger et al. 2009; Shi, Han et al. 2010; Du, Chen et al. 2011; Abdelkader, Cooper et al. 2012). It worked successfully in most of the cases. However, in practice, the parameters like the window size and the optimal rank of the local multivariate regression need to be optimized carefully. Problems occur when PDS estimates the rank of the local PCR/PLS models developed on the windowed data. Bad local rank estimations could create artifacts in the transferred spectra. A procedure was proposed to detect and reduce PDS artifacts by examining the spectrum of an independent sample before and after standardization (Bouveresse and Massart 1996). Preprocessing before standardization has been proved beneficial to reduce noise. Data decomposition as PCA also helps to reduce these PCs dominated by noise effect to filter noise before standardization by PDS. Moreover, PDS is not a preferable method to be employed for process monitoring and control, since a different calibration model for the same constituent at each incidence of maintenance.

5) Wavelet Transformation before standardization

Standardization can also be applied to model differences between spectra transformed to another domain. Wavelet transform (WT) enables the time-frequency representation of the instrumental signals. A wavelet ψ is a function of zero average:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{0.43}$$

which is dilated with a scale parameter s, and translated by u:

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \tag{0.44}$$

The wavelet transform of f at the scale $\boldsymbol{s}$ and position $\boldsymbol{u}$ is computed by correlating f with a wavelet atom. Wavelet coefficients can be expressed as follow:

$$Wf\left(u,s\right) = \int_{-\infty}^{+\infty} f(t)\frac{1}{s}\psi^{*}\left(\frac{t-u}{s}\right)dt \tag{0.45}$$

Wavelet coefficients have information on both frequency and position, while Fourier coefficients only refer to the frequency.

The original signal is passed through a low-pass filter (known as scaling filter, **H**) and a high-pass filter (called wavelet filter, **G**). These filters are orthogonal. WT analysis decomposes a signal function (for example, a spectrum of a sample) into a set of outputs as approximations (low-frequency components) and details (high-frequency components) at different scales (levels) and positions. The outputs of both filters at level 1 are composed of a set of N/2 coefficients of approximations and N/2 coefficients of details, where N is the length of the signal. The decomposition is an iterative process on the low frequency components only and continues till the set consists of a single unit. This process is known as Mallat's pyramid algorithm (Mallat 2009) and completely recursive. The main advantages of WT are associated with signal compression and denoising. Application of WT-related methods have been studied extensively in spectroscopic signal processing and plays an important role in both NIR and IR spectroscopy (Chau, Liang et al. 2004). Compared with Fourier transform, the conventional analyzing tool in signal process, wavelet transform are more efficient than Fourier at compressing near infrared spectra (Fearn and Davies 2003). Wavelet analysis could also be applied to correct constant or non-constant background (Tan and Brown 2002). Compared with other baseline correction methods such as first derivative, multiplicative signal correction (MSC), orthogonal signal correction (OSC) and the polynomial background approximation method, it was demonstrated to be an efficient method for removing the non-constant background variation automatically for both simulated and real NIR data. Igne et al. applied wavelet and Fourier transform to soybean protein and oil calibrations for comparison (Igne and Hurburgh 2010). Both intra-brand and inter-brand standardization were examined by different frequency filtering methods. The results showed the possibility of calibration transfer by modifying the signal before calibration to avoid the use of standardization methods.

Walczak et al. first proposed a method for comparing the performance between two NIR instruments in the wavelet domain (Walczak, Bouveresse et al. 1997). They applied WT to the spectra of a subset of standardization samples obtained on both master and secondary instruments. Then relate the WT coefficients of the NIR spectra obtained from the two different instruments utilizing univariate linear models. This model was set up to determine the standardization parameters between the two sets of NIR spectra. Once these parameters were computed, the inverse wavelet transform was performed with filter n applied to the transferred wavelet coefficients and obtain standardized NIR spectra. The signal (spectral variation) is compressed by discrete wavelet transform, which enables greater stability in the transfer. In this research, the results of soy samples indicated that when there was a global nature in the spectral differences between the master and secondary instruments, advanced standardization methods like PDS or SWD do not perform better than the simple slope/bias correction method. However, both PDS and SWD were applied to a small amount of standardization samples. The results showed RMSEP reduced significantly when the number of standardization samples was greater than 5. The wavelet transform could also be performed on scores or factors in principal component analysis and partial least squares, which was proposed as indirect standardization method (Park, Ko et al. 2001). In this method, only a few transfer coefficients were taken with similar performance to the spectrum transfer case. Those two methods both utilized univariate standardization models, which limited the information of calibration transfer. By using univariate models instead of PDS for standardization in the wavelet domain, it lost partially the benefits from conventional PDS. Moreover, although WT has noise-filter effect, it still did not achieve full robustness to spectral noise.

A modified standardization method called wavelet hybrid direct standardization (WHDS) was proposed by Tan et al. (Tan and Brown 2001). There are two main differences between WHDS and standardization in wavelet domain (SWD). First, the reconstruction is made before the standardization in WHDS. In this new wavelet reconstruction algorithm, approximation and detail spectra are reconstructed separately. Second, piecewise direct standardization (PDS) and direct standardization (DS) were used to correct the differences of original spectra by transforming the reconstructed approximation and detail spectra, respectively, rather than a simple linear regression in SWD. This method was proved to be useful to remove baseline structure that varies between the calibration and test sets. In this research, the proposed method was applied to NIR data, and its performance was compared with that using conventional methods. The results indicated that combining PDS and DS with the wavelet multi-resolution technique, WHDS algorithm allows a more robust and reliable means for standardization as its robustness to high-frequency noise. The effect of the number of standardization sample sets was also discussed in this research, which would be the subject of chapter 3 in this dissertation.

Wavelet packet transform (WPT) derives from WT (Walczak and Massart 1997). WPT allows a full multi-resolution analysis and d_j (reflects the high-frequency information contained in the original data set at the *j*th level) is also involved at the same time to produce the scale and wavelet coefficients at the next resolution level. The discrete WT (DWT) is generalized in the WPT procedure to provide a more flexible tool for analytical data analysis.

6) Neural Network (NN)

Nowadays, ANN is commonly known as a non-linear multivariate calibration method, although satisfactory results could also be obtained if the unknown relationship is actually linear. It has been applied to many situations that involve problems of prediction, classification or control. ANN has great ability to construct a relationship between two sets of variables as spectra collected from two spectrometers. ANN possesses the flexibility due to the distribution of information among multiple weighted connections and nonlinear processing units of the network. On the other hand, this flexibility can become a weakness if these parameters are not optimized. ANN is particularly subject to over-fitting depends on the setting of parameters. Therefore, the error goal of the training of neural networks should be carefully chosen to avoid both under and over training.

When it comes to instrument standardization, the task is to model output variables from input variables in order to correct spectral differences. The problem of using ANN for standardization is the number of standardization samples is usually not large enough to estimate the parameters in ANN modes, which leads to over-fitting problems.

Without enough samples in training set, Duponchel et al. (Duponchel, Ruckebusch et al. 1999) tried to use ANN for standardization by containing as many input neurons as wavelengths on the secondary instrument spectra and as many output neurons as wavelengths on the master instrument spectra. To solve this problem, the number of links was reduced and a moving window based method was created to relate to a window on the secondary instrument. In this study, a multilayer perceptron, with forward propagation of activation and back propagation of error was implemented for the correction of the spectral data. Compared with other conventional standardization methods as the patented Shenk algorithm, DS and PDS, ANN gave slightly better results than PDS and showed the smallest SEPs. Although, the small differences in SEP between ANN and PDS may not be significant, ANN showed its correction ability by comparing the difference spectra before and after the correction. However, with this method, ANN models were built as many as the spectral window and many parameters to be optimized through the experimental design. It is worth to note the importance of the optimization of adjustable parameters. Neural network behaviors are very sensitive to the error defined to stop the training and avoid over-fitting.

In order to resolve the inharmonious facts between the requirement of a large number of training set and a low number of standardization samples, Despagne et al. (Despagne, Walczak et al. 1998) proposed to truncate spectra in finite-size windows and assess a position index to each window. With the similar window based concept, each spectral window was considered as a separate standardization sample and one global model was built for all spectral windows. The procedure to avoid local over-fitting was to determine the number of epochs by an independent monitoring set, which can be built with all spectral windows of a representative sample. New external samples collected on the secondary instrument were corrected by the connection weight of the monitoring set. No additional background correction was needed in contrast to PDS due to the ability of model different types of baselines in NNs. Six different secondary instruments were used in this study for standardization using both NN and PDS for comparison. The effect of the subset selection by five different methods was also examined, which indicated that the NN approach is extremely sensitive to the standardization samples selected. This would be further discussed in the section of standardization samples below. In this study, calibration results on transferred spectra (RMSEP) were suggested to be an evaluation of the quality of standardization instead of the transfer residuals. The reason was that the authors illustrated the possibility of the amplitude of differences between two instruments caused by background noise (stray light) rather than the differences on a calibration models. The calibration results displayed better performance of NNs than PDS. However, the authors claimed both NNs and PDS had their merits, while NNs performs better when there are stray light, cell path-length and wavelength changes.

Moreover, studies were conducted to combine WT and ANN (Aminian and Aminian 2000; Chalus, Walter et al. 2007) for a data compression purpose. The wavelet coefficients obtained by decomposition were used as input variables for constructing ANN models with limited numbers. Another incorporation is a feed forward neural network based on wavelet transform called wavelet neural network (WNN) (Balabin, Safieva et al. 2008). WNN uses wavelet functions instead of the traditional sigmoid function as its transfer function in each neuron. Therefore, this network poses the advantages of both the wavelet transform in de-noising, background reduction and recovery of characteristic information and the flexibility of neural network.

7) Orthogonal Signal Correction (OSC)

Originally developed as a method for spectral preprocessing, it was proposed by Wold et al. (Wold, Antti et al. 1998) to remove systematic noise such as baseline shift and multiplicative scatter effects without eliminating any information related to **Y** matrix (reference information). The objective of this method is to calculate a weight vector **w** that satisfies

$$\mathbf{t} = \mathbf{X}\mathbf{w} \tag{0.46}$$

to be orthogonal to **Y**, where **t** refers the PC scores. The corrected score matrix **T** is then used to construct the PLS models. Compared with other preprocessing methods such as MSC, SNV, OSC tends to use less PLS components and complexity in calibrations (Blanco, Coello et al. 2001).

An alternative algorithm was proposed (Fearn 2000) to substract from **X** factors that gives rise to increase the correlations between the scores and **Y**. When applied to the training data set, the first *f* factors are removed from **X**:

$$\mathbf{X_0} = \mathbf{X} - \sum_{i=1}^{f} \mathbf{t}_i \mathbf{p}_i^T \tag{0.47}$$

where $\mathbf{p_i} = \mathbf{X^T t_i} / \left(\mathbf{t_i^T t_i}\right)$ denotes the vectors of loadings. The loading weights are applied to $\mathbf{X_0}$ not **X**, so the scores will be more strongly correlated with **Y** due to the removal of the unrelated information.

OSC has also been applied to calibration transfer between instruments (Sjöblom, Svensson et al. 1998). For standardization, OSC is aimed to exclude unrelated vectors to the factor of interest **y** vector. After preprocessed by OSC, spectra could be more transferable due to the removal of the instrument signatures and less dependent on instrumental variations. In this algorithm, PC score vector **t** was first centered and then orthogonalized to **y**:

$$\mathbf{t}_{\perp} = \mathbf{t} - \frac{\mathbf{y}'\mathbf{t}}{\mathbf{y}'\mathbf{y}}\mathbf{y} \qquad (0.48)$$

Then add the average **t**:

$$\mathbf{t}^{*} = \mathbf{t}_{\wedge} + \mathbf{t}\bar{\mathbf{t}} \qquad (0.49)$$

The loading vector **w** is calculated by

$$\mathbf{w} = \mathbf{t}^{*'}\mathbf{X} / (\mathbf{t}^{*'}\mathbf{t}^{*}) \qquad (0.50)$$

Then scale it to the unit length:

$$\mathbf{w} = \mathbf{w} / \sqrt{\mathbf{w}'\mathbf{w}} \qquad (0.51)$$

The last step is to create a new $\mathbf{t}$ vector from vector $\mathbf{w}$ using equation (2.29). It is notable that the slight difference between this algorithm and the one in the previous research (Wold, Antti et al. 1998) is this is an iterative procedure for finding a vector of scores $\mathbf{t}$. Calculation repeats until $\mathbf{t}^{*}$ becomes stable. New spectra were corrected by removing the OSC-component by using the loading $\mathbf{p}$

$$\mathbf{X}_{\mathbf{OSC}} = \mathbf{X} - \mathbf{t}\mathbf{p}' \qquad (0.52)$$

In comparison with other signal correction methods before calibration, OSC modeled with nearly 100% variance in the first component. By including corrected spectra of training set obtained on both instruments, a PLS model was built to predict corrected spectra of prediction set collected on both instruments. The results showed lower RMSEP. This might be because that by including samples from both instruments, the PLS-model tends to compensate for the instrumental variations. The problem here is that samples need to be measured on both instruments. Another factor of effect that is worth to note is the number of OSC components. In this research, two OSC-components were required to make a difference from the MSC treated model. However, excessive use of the OSC components would lead to identical spectra and over-fitting.

Orthogonal projections to latent structures (O-PLS) was proposed as a preprocessing method to remove variations from X that are uncorrelated to Y , which is the same goal of OSC (Trygg and Wold 2002). Instead of removing the orthogonal signal from the PC score matrix (**t**) prior to calibration development, the O-PLS method removes the orthogonal information from the PCs calculated by PLS. Then develop a calibration model on the reconstructed X matrix. O-PLS should have some advantages over OSC such as the improvement on outlier detection and calculation efficiency due to no time-consuming internal iteration process. The results of the comparison between an original PLS model and PLS model built on O-PLS matrix showed better interpretation of PLS models, less complexity and prediction performances. As an effective preprocessing method, O-PLS was proposed to be combined with ANN (Wang, Liu et al. 2009) for input dataset. O-PLS treated data was used as input variables for ANN models instead of the original data set. Compared with first and second derivative spectra, O-PLS pretreated spectra lead to calibration models with the best prediction performance and more efficient in calculation.

Based on the theory of OSC, a series of related versions were developed for spectral correction before standardization. In the area of calibration transfer, a method called transfer by orthogonal projection (TOP) was proposed for deriving robust and transferable calibrations for various instruments (Andrew and Fearn 2004). The scheme of this method is to orthogonalize the spectra to directions that are most related to instrumental variations. It is the information on between-instrument variations that was used to remove this variation from the spectral data. A small number of standardization samples were scanned on each instrument. Then average over samples to obtain a matrix R and PCA was performed on this matrix to get the directions of instrumental variations. Orthogonalize X to the PC score matrix and the projection of *X* is calculated by:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{X}\mathbf{P}\mathbf{P}^{\mathbf{T}} \tag{0.53}$$

where P is the matrix of orthonormal eigenvectors of R. Then, $\tilde{\mathbf{X}}$ was used to construct the calibration models. Compared with the application of OSC on calibration transfer, TOP uses the extra information (instrumental variations) from standardization set to remove exactly those dimensions that interfere with the transfer. TOP requires a standardization set of samples measured on multiple instruments, which may not be realistic in practice. With the same approach applied on eliminating variations caused by temperature, external parameter othogonalisation of PLS (EPO-PLS) (Roger, Chauchard et al. 2003) was suggested to cope with cases with immeasurable variations like on-line analysis. This research gave a broader scope of this method for many other applications such as making calibrations robust to humidity or path-length variations.

Thereby, a new method called dynamic orthogonal projection (DOP) (Zeaiter, Roger et al. 2006) was developed to improve the robustness of NIR calibration for on-line applications. Since there are so many unknown influence factors (most are unexpected) during the on-line process, it was expected to compute an orthogonal correction on the original spectral data of some virtual standards created by a kernel function. In calibration transfer, virtual standards spectra were created by estimating X_τ as a linear combination of X_0:

$$\mathbf{\hat{X}_\tau = AX_0} \; with \; a_{ij} = Fy_{\tau i}(y_{0j}) \tag{0.54}$$

where $Fy_{\tau i}$ is a kernel function centered on $y_{\tau i}$. Then a difference matrix $D = X_\tau - \hat{X}_\tau$ was calculated, which was used to estimate an orthogonal basis P (similar to TOP). The reconstructed X was calculated by equation (0.55) and used for calibration. In this case, this method should be able to handle different kinds of perturbations. DOP does not require standardization samples to be scanned on multiple instruments and gave rise to more flexibility for on-line applications.

Among these orthogonal projection methods, OSC and O-PLS are the first kind of approaches that remove factors unrelated to **Y** matrix from **X** matrix. Loadings of these factors are substracted from the **X** matrix before calibration. Every standardization sample need to be scanned on multiple instruments and orthogonalized. The other kind of orthogonal method includes TOP, EPO-PLS and DOP, which estimate the external parameters and removes them from the **X** matrix. In this case, only a set of standards are needed to estimate the differences caused by external factors.

These orthogonal methods as well as their modifications and extensions were compared for their performance of standardization in a case study of whole soybean protein and oil models (Igne, Roger et al. 2009). A network of four instruments from two brands was tested for inter- and intra-brand calibration transfer by using orthogonal projection techniques. Whole soybean samples collected from four crop years were used as calibration set, and two distinctive validation sets included 20 samples selected over years and 40 samples from the latest year. A set of 20 standardization samples was created by randomly selected from the calibration set just to check the concentration range that covered the remaining samples. PDS, slope and bias and exhaustive calibration models (included spectral data from two different brands of instruments) were used to compare with the transfer performance of these orthogonal methods. Calibration models built on spectral data of its own were used as benchmark for comparison. As the results for intra-brand transfer, there were no significant differences among the results obtained by different methods. Many of them reached better precision than their original models. The improved performance on the secondary instruments demonstrated the benefits of removing unrelated information of the instrument variations. As for the inter-brand transfer, the performance between the two master instruments showed no significant results on validation set 1. TOP and EROS provided similar results to the original model with other methods significantly less precise. In comparison with these classical standardization methods for intra-brand calibration transfer, the best performance obtained by orthogonal method was OSC, which is comparable to the exhaustive calibration models and simple slope and bias correction. As the authors state, the choice of transfer samples might be responsible for the present situation. Thereby, the best option might be choosing standardization samples that optimize orthogonal methods.

1.4.7 Selection of standardization samples

Most multivariate calibration methods for standardization require a set of standardization samples to be measured on multiple instruments. In order to establish the transfer parameters for standardization, the standardization samples need to be representative to the entire experimental domain and stable enough between the situations in which the standardization is performed. (Burns and Ciurczak 2007). The two main approaches for selecting standardization samples are selection of a representative subset of standardization samples from the original pool and the use of independent standardization samples. For unstable samples, independent standardization samples are needed for standardization (Park, Ko et al. 2001). Selected representative subsets were used for calibration transfer in this research given the characteristics of the NIR network and agricultural samples we worked on.

Approaches for sample selection have been reviewed in the previous section. These approaches are mostly discussed on how to partition the training and test sets (Ferré and Rius 1996; Daszykowski, Walczak et al. 2002; Galvao, Araujo et al. 2005). However, little attention has been paid to select the proper standardization samples for calibration transfer. As the set of standardization samples does impact on the performance of calibration transfer, the choice of standardization set need to be studied associated with the standardization methods applied. Unlike dividing samples into training and test set, the essence of standardization samples is to represent the instrumental variations, while other source of differences caused by external factors should be minimized. Thereby, in order to be scanned on multiple instruments, samples need to be stable and less subject to variation and error which are not interest within the calibration model. With these constrains, different kinds of samples could be used for standardization. If samples are considered to be stable under regular conditions over a period of time, then the standardization set could be selected from the original data pool. For those samples that are hard to be stored, generic standards or resembled chemicals are options for standardization set.

To test the effects of different standardization samples on calibration transfer, Bouveresse et al. (Bouveresse, Massart et al. 1994) conducted experiment on three different sets of samples: samples in calibration set which were similar to agricultural samples, generic standards and the mixture of pure organic and inorganic chemicals. Sample spectra measured on the secondary instrument was transferred to the primary system using classic Shenk's patent method (Shenk and Westerhaus 1989). The first set of 30 agronomic samples showed the best results of standardization. Even the second and the third sets had larger optical density than the first set, they did not perform better than the first one. The shift of the mean spectra between different instruments of the first set showed its smaller instrumental differences than other sample sets. This attracted our attention to the presence of systematic bias that proved the importance of the selection of standardization samples. The systematic bias (constant shift) is independent to the wavelength and could be removed by the derivative spectra. Several hypotheses were proposed to explain the presence of the bias, which were all related to external variations such as the change of products over time, contributions of the cells and the temperature fluctuation. This research indicated that the standardization samples should be of the same nature as the prediction set, and be measured under the same condition including sample presentation, temperature and simultaneousness as much as possible.

Two different types of standardization samples could be either selected from the calibration set or prediction set. Selecting samples from the calibration process, standardization samples would have the same nature as the calibration samples. Different sample selection approaches are applied in this step during calibration transfer process. Choosing samples with the highest leverage (Wang, Veltkamp et al. 1991) was employed to select standardization samples when the authors firstly proposed piecewise direct standardization. Moreover, Kennard-Stone algorithm(Kennard and Stone 1969) is a commonly used selection method for selecting standardization samples.

The criterion for a successful standardization is for the average RMS(C) for the test samples between each pair of instruments to be equal to or less than that between re-packs on the same instrument (Dardenne and Biston 1991). It was demonstrated improvements were obtained with less selected representative samples in the subset using PDS (Bouveresse and Massart 1996). In this research, the influence of the subset size and three selection methods were examined. The results showed better performance of the subsets with representative samples that covered the entire experimental space. Better standardizations were achieved with increasing number of standardization samples regardless of the selection methods.

To select the most representative subset for calibration standardization, several algorithms were compared along with specific calibration transfer methods. Siano et al. used three algorithms (Kennard- Stone, Leverage and OptiSim (Clark 1997)) to select representative subsets for standardization (Siano and Goicoechea 2007). RMSEP decreased first then increased with increasing numbers of selected samples. The subset with 20 samples selected by OptiSim obtained the lowest RMSEP for calibration transfer using PDS. This study also compared the transfer performance of three different standardization methods: direct standardization (DS), piecewise direct standardization (PDS) and wavelet hybrid direct standardization (WHDS). Using the optimal subset selected by OptiSim, WHDS improved the secondary spectra after standardization significantly.

With the purpose of reducing the number of samples, the comparative analysis using restructured near infrared and constituent data (CARNAC) was applied for quantitative analysis of biscuit dough (Davies and Fearn 2006). In this method, it is assumed samples with the same spectra should have the same composition. The assumption stands when the database is large and possibly contains all variations of prediction samples. By a selection procedure, samples that are similar to the prediction samples are selected from the database. Data compression and modification are required to stress the factor of interest. The best RMSEP was obtained by the optimized program. The limitation of this type of selection methods referred as "local" method including CARNAC, LWR and LOCAL (Pérez-Marín, Garrido-Varo et al. 2005), is the original database need to be large enough to accumulate variations over time and space, which is not always possible in practice.

Roussel et al. (Roussel, Hardy et al. 2001) used different calibration sets for calibration transfer of genetically modified grains (GMO). With a larger database (8180 scans), the accuracy of classification was improved by at least 10 percentage points for PLS models and 6 percentage points for ANN models. The results showed no improvement of accuracy reached by SNV or PDS, which indicated more inclusive standardization set would be needed.

Even using the same Kennard- Stone algorithm for sample selection, different subsets could be obtained by whether starting the first point from the center of the experimental space. Compared to PDS, neural networks (NNs) is extremely sensitive to the selection of standardization sets (Despagne, Walczak et al. 1998). For standardization with NNs, it is suggested to identify and include the central point first, then add each new sample to maximize the Euclidean distance when using KS.

Selecting samples from the prediction set extends the representativity of the standardization set with new source of variability from the prediction set (Bouveresse and Massart 1996). However, this approach would only apply to the situation when the prediction samples are available before building the calibration models. Because standardization samples are needed to be measured under the same condition as the calibration samples. Thereby, it could not be applied to calibration update.

A diagnostic tool (Jouan-Rimbaud, Massart et al. 1997) was developed for determination of representatives of standardization and calibration data. Two statistical methods were proposed. One is to compare the variance-covariance matrices to determine whether the two data sets have similar volumes both in magnitude and direction; the other is to compare the data set centroids by computing the Mahalanobis distance between the means of each data set. Another simple and straight way could be examine the direction of PC scores.

Agricultural products are very complicated substances with oil, protein, moisture and carbohydrates present in various combinations. They are not perfectly stable over several years, but they can be used as standardization samples due to their stability over the space of a certain amount of time under a proper storage conditions (Burns and Ciurczak 2007). Thereby, the main focus of this study is selection of subset samples from a large population accumulated over the past ten years.

1.4.8 Number of standardization samples

There is no criterion to define the exact number of standardization samples needed to perform a calibration transfer successfully. The optimal number of standardization samples varies case by case in practice. However, the set of standardization samples does make a difference during the calibration transfer process.

In a more recent study, a calibration transfer method called spectral regression (SR) was proposed (Peng, Peng et al. 2011). With comparison to the classic standardization method PDS, spectral regression method created the subspace variables to associate with the transformation matrix. The effect of the number of subset samples on different calibration transfer methods was also evaluated by examining subsets selected using KS algorithm. While the performance of PDS was less sensitive to the change of numbers in standardization sets, SR gave better result when standardization samples were sufficient.

1.4.9 Temperature standardization

In agricultural and food products, water is an essential content, which has an important impact on the NIR spectra due to its absorptions. Also, the NIR spectrum of water is very sensitive to temperature. This is because the weak forces influence the molecular bonds such as hydrogen bonding, which could be mostly affected by the change of temperature. Moreover, the overlapping bands at 1450 and 1930 nm of water bands only make the changes more complicated. Details on the spectral changes and hydrogen bonds could be found in reference (Maeda, Ozaki et al. 1995). The effects of hydrogen bond and sample temperature were examined for the reliability of NIRS results (Büning-Pfaue 2003).

In NIR spectra, a temperature raise will increase the probability of molecules transfer to a high energy level, which means more radiation is absorbed and less is reflected. The increased sample temperature raises the vibration energy between the molecules so that molecular bonds, especially the hydrogen bonds break. Accordingly, the clusters of the water molecules become smaller and the absorbance of the free hydrogen bonds increases. In correspondence, the reflectance decreases, while the absorbance increases with increasing temperature. In Figure 7, soybean spectra scanned at three different temperatures (5, 22 and 45°C) were plotted. Increasing absorbance at the water peak at 970-980 nm wavelength regions was observed with higher temperature.

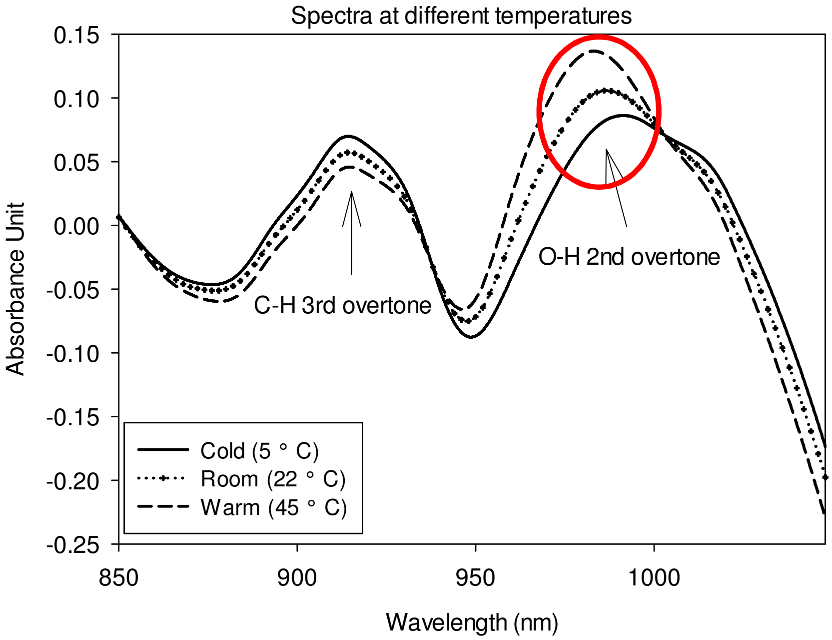


Figure 7. Soybean spectra taken at different temperatures

It has been found in the early 1980s that the effects of temperature on NIR protein are responsible for the differences discovered between Federal Grain Inspection Service (USDA-FGIS) Laboratories. Williams et al. (Williams, Norris et al. 1982) developed a method to correct the temperature of grounded wheat samples to room temperature. The calibration model was built on samples from all temperature ranges. The authors indicated that the temperature effect was not associated with significant slope change of the calibration equations. Thus, it could be corrected by means of intercept adjustment. An inverse relationship was found between temperatures of ground samples and their corresponding apparent protein contents. Several methods were suggested to compensate for the differences caused by temperature fluctuation. One method is to select temperature – independent wavelengths that are least sensitive to temperature fluctuations (Wulfert, Kok et al. 2000). A second approach is to analyze samples at the same temperature at which the calibration was developed. This is hard to implement in reality when samples are not scanned at the same time. The third suggestion is to develop a temperature compensation factor using a correlation chart (Kawano, Abe et al. 1995). An approach called K-method was developed based on a sample combination to reduce the bias caused by changes of temperature with calibration equation. The applicability and universality were demonstrated on chemical components of sucrose, fructose and glucose (Abe, Iyo et al. 2000). Selected wavelengths were used to estimate the effect of the deviation of sample temperature on the bias. Bias could be reduced by selecting adequate wavelengths for calibration, which could not always achieved by MLR. This study gave mathematical and theoretical analysis of K-method and proved its applicability. However, the situation of spectra change due to temperature fluctuation in real agricultural products is more complex than pure or mixture of chemical solution, which needs further study.

Temperature compensation techniques are mainly discussed in this study. NIR calibration procedure involves developing a base calibration followed by the addition of samples to the base calibration for instrument stabilization and temperature stabilization. The same procedure could be used to validate the stability of existing calibrations to differences in sample temperature. Instrument temperature is usually controlled by heating and cooling circuitry within an instrument. The goal is to stabilize sample temperature effects in new NIRS calibrations and/or to test the stability of calibrations to sample temperature. The effect of temperature on spectral information could be visualized in the PC scores plot (Figure 8). Same temperature samples were scanned at different temperatures. Cold and warm samples tend to have its own clusters, whereas cold samples on the right with positive scores on PC 1 and warm samples have negative ones.

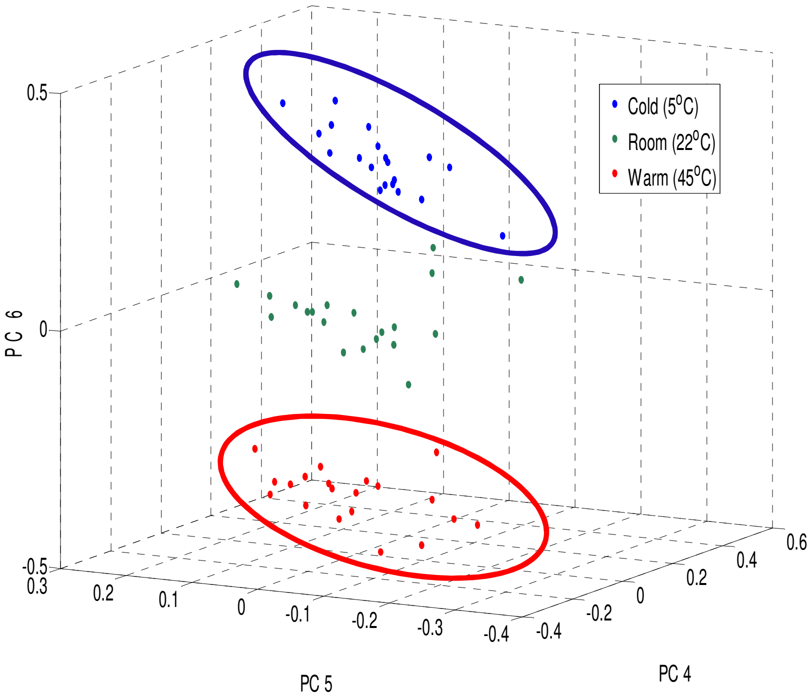


Figure 8. Principal component score plot on preprocessed soybean spectra (second derivative, SNV and DT) scanned at different temperatures

Wang et al. (Wang and Kowalski 1993) demonstrated that with some modification of PDS, it is possible to standardize between discrete responses like temperature. The differences of corn spectra collected under different temperatures were compensated after standardization, which was far large than the sample-to-sample variations without standardization. Moreover, with more standardization samples (9 compared to 4) included, the differences among spectra were further reduced. Significant improvement was found in SEP of content B due to information recovers by temperature correction. However, for grain products, there are possibilities of experiencing cold and hot temperature depends on the weather during handling process in North America. The cold condition, sometimes frozen samples need to be examined too.

It was observed that the effect of temperature fluctuation on NIR reflectance spectrum is non-linear (Smith and Gemperline 2000). Peirs et al. (Peirs, Scheerlinck et al. 2003) studied the temperature effect on spectral data and calibration models were examined on apple fruit soluble solids contents. The experiments showed the temperature have influences on the pattern of the spectra, which was useful for predicting the temperature of an individual samples. Two kinds of models were built: temperature dedicated models (built on samples collected at each temperature separately; global temperature models (include samples collected at all different temperatures. As for the effect of temperature-induced spectral variations, prediction performance of calibration models was assessed. Dedicated temperature calibration models were sensitive to sample temperature deviations. It is less accurate to validate a calibration model at a different temperature as it was developed than interpolated results. On the other hand, the global temperature models yielded more accurate results, which were even better than the calibration models developed at the same temperature of the validation set. Building one global temperature model is more practical, which treats temperature as an unknown variable. This enables the global model robust to temperature changes and certain measurement deviations for future prediction. Another advantage of the global model is it is not necessary to know the exact values of the temperature, which is the case in most common case. The only problem is the unknown deviation of prediction temperature from the calibration temperature. There would be no universal solution, but only case by case studies on sample properties, handling conditions and any other environmental factors.

The global robust temperature calibration model was also built on protein and fat contents for a complex food model system (Zhang, Chang et al. 2010). This study focused on the optical mechanism of temperature effect on spectral measurement and its adverse effect on calibration models. In terms of optical properties, no distinct variation on absorption coefficient was observed, while scattering coefficient decreased with increasing temperatures in the whole spectral region. This is due to the decrease in diameter and volume fraction of fat globules that reduced the scattering. Also, with a wide melting range, fat fully melted at 40 ℃. Thus, the state of fat changed that lead to a significant decrease on scattering coefficient as the temperature gets near to the melting point. To correct the temperature effects on optical measurement, the authors conducted an experimental design associate with the fat and protein concentrations of raw milk to ensure the uniform distribution with a wide coverage and independence between fat and protein constituents. This was achieved by adding whey protein concentrate and protein powder and oil mixture to adjust their proportions. The global temperature models were established on spectra obtained at different temperature and compared to four local models built at one specific temperature. The local temperature models were demonstrated to be sensitive to temperature changes in terms of higher RMSEP compared to the global temperature models. When the exhaustive calibration sample set covers all the variations both of the response of interest but also of sample temperature, a global temperature model is suggested to improve the calibration robustness to temperature fluctuation. The difficulty for its application is that design of experiment is not always possible for agricultural products.

External Parameter Orthogonalisation of PLS (EPO-PLS) was introduced as an standardization method in previous section (Roger, Chauchard et al. 2003). EPO was presented as a preprocessing method that removes the information in spectra that are mostly influenced by the external parameter variations as temperature of products or spectrometer. Meanwhile, it also reduces the space dimensionality with regard to external parameters. There are two ways that could serve this purpose; one is to find the subspace that orthogonal to Y, the other is to find the subspace in which the influence of external factors occurs. EPO belongs to the second way. Experiment was conducted on measurement of sugar content of apple. Results of prediction bias obtained by model developed at 25 ℃ were linearly varied with temperature levels without any correction. EPO preprocessing reduced RMSEP significantly with optimal parameters selected for the algorithm. The advantages of this method are only a small set of appropriate samples are required to be measured at different temperature levels and its feasible application to any existing calibrations.

Various paper described temperature compensation methods and their application on different samples and systems. However, the effect of temperature on the accuracy of protein and oil measurements of grains bases on NIR transmittance spectroscopy has not been addressed so far.

1.4.10 Noise robustness

Differences among replicate measurements of the same sample spectra reflect the within sample variability sources of the NIR spectra. This could be ascribed to a high-frequency component usually known as the instrumental noise and a low-frequency component due to the differences in the specific nature of samples (Foca, Ferrari et al. 2011). Several papers in the literature suggest different strategies for denoising of NIR spectra in terms of instrumental noise, there is no previous knowledge regarding the low-frequency component of variance signals associated with sample properties.

Many researches have been conducted to minimize instrumental noise and its effect on calibration models. Several spectral preprocessing techniques are developed to serve this purpose. Study shows that standard normal variate (SNV) preprocessing was very efficient in improving robustness for certain types of noise (Roussel, Igne et al. 2011). SNV eliminated multiplicative and baseline shift noise as well as wavelength shifts and improved model robustness for all calibrations developed by different regression methods.

To build a calibration model against noise, strategies were developed by adding noise simulation to spectral data (Sáiz-Abajo, Mevik et al. 2005). Spectral noise could be attributed to external perturbations such as temperature fluctuation that may cause a systematic error and instrumental difference collected at different session that lead to random errors. Six different noise simulations were applied to spectra of wine samples. The proposed ensemble method and data augmentation process improved the predictive ability of model significantly with reduced number of samples included. Similar data augmentation method was applied in pharmaceutical research for active ingredient prediction (Pieters, Saeys et al. 2013). In practical application, a large amount of variations come from various batches that do not correspond to the factor of interest. Artificial noise augmentation was obtained by adding the mean-centred spectra of unknown chemical variation to *n* repetitions of the original calibration spectra. Prediction performance was improved with reduced variability in calibration set.

1.4.11 Robust calibration models

Building a global robust model has been a common approach to incorporate external variations that are not the factor of interests in the calibration model. In recent years, compared to building a global robust model, techniques based on variable selection are also proposed to develop robust calibration models with the presence of external variations. The essence is to select spectral variables that are insensitive to external variations, which has already been applied to deal with temperature fluctuations. In this case, spectral regions that are related to factor of interest but highly involved with external influence, not linearly or indirectly related to factor of interest are excluded before calibration procedure. Various variable selection methods are developed to serve this purpose, such as simulated annealing (SA) (Kalivas, Roberts et al. 1989), genetic algorithm (GA) (Lucasius, Beckers et al. 1994), interval partial least squares (iPLS) (Norgaard, Saudland et al. 2000), uninformative variable elimination (UVE-PLS) (Cai, Li et al. 2008), and newly proposed method for key wavelength selection using competitive adaptive reweighted sampling technique (CARS) (Li, Liang et al. 2009). It is worth to note that since the number of variables is significantly reduced after selection, the risk of over-fitting should be avoid due to degree of freedom lost in the selection.

Basic concepts of robust techniques were reviewed in Ref. (Daszykowski, Kaczmarek et al. 2007). Those fundamental concepts of robust statistics help one to optimize the calibration procedures including outlier detection, examination of covariance and correlation matrices and leverage and Mahalanobis distance, all of which enhance the understanding of the experimental data before calibration models are actually constructed.

Research (Swierenga, Wülfert et al. 2000) using simulated annealing (SA) for variable selection was conducted to compare this technique with global models in the presence of continuous temperature change. Separate models were built using calibration samples measured at various temperatures. With only 10 and 25 wavelength points selected by SA for the mixture of chemical samples and density of high oil products respectively, RMSEP were comparable or even better than the global models using the whole region of 200 data points. Results of using SA were also compared with that of UVE-PLS, eliminates variables from PLS models by judging a criterion based on regression vector. SA variable selection model performed significantly better than the UVE-PLS model with respect to the prediction error at different temperatures. The disadvantage of SA is that it requires special expertise to tune parameters as PLS factors, the number of variables selected and length of Markov chain to obtain optimal prediction, which could be less accessible than global models for commercial application. NIR calibration robustness was examined on the internal quality attributes of mandarin fruit as its total soluble solids (TSS) and dry matter (DM) (Guthrie, Reid et al. 2005). Robustness was tested across changes in terms of season, locations and time.

Zeaiter et al. (Zeaiter, Roger et al. 2004) reviewed different definitions of the concept of robustness and methods that are applied for evaluation, especially with the application of NIR in industry. The authors adopted the definition of robustness of a multivariate calibration model as "the stability of its predictive capacity against perturbations centered on standard conditions". In the literatures, robustness is often associated with terms as accuracy, precision and uncertainty. The accuracy of the test of NIRS calibration is composed of the total variance defined by SEP:

$$SEP^2 = S_r^2 + S_{NIR}^2 + S_e^2 \tag{0.56}$$

Where S_r is the repeatability of the reference method, S_{NIR} the repeatability of the NIRS method and S_e is the lack of fit of the calibration model. True accuracy is unreachable, which requires removal of error in reference method, usually wet chemistry method in NIR models. But it could be estimated by a bias value observed from long term operations. The term of precision describes random errors and usually is expressed as variance and standard deviation. It is indicated that good precision does not assure good accuracy in the presence of significant systematic errors. Uncertainty is related to accuracy that also describes the closeness of the values of experimental means to the true values. Uncertainty focuses on the range of the values evaluated from the statistical distribution of the results, but is not able to be used for correction of the results as errors.

Strategies for robustness assessment involve the use of a robustness test, which is defined as “An experimental study in which one evaluates the influence of small changes in the operating or environmental conditions on measured or calibrated responses”. The changes include different external conditions and instruments. Major steps of the robustness test (Swierenga, de Weijer et al. 1999) are selection of a representative subset of samples, external factors, factor levels, and experimental designs, prediction of the desired sample parameters and computation of the effects of the external factors on the model prediction error. Aimed at minimizing the objective function of prediction error, robust indices are used for robustness assessment, which include signal-to-noise ratio (SNR), response surface methodology, the robustness coefficient (RC) and root mean square error of prediction (RMSEP).

It has been mentioned that preprocessing techniques provide improvement on calibration robustness by removing spectral variation not due to the parameter of interest as well as enhancing the wavelength selection for the parameter of interest. The robustness of multivariate calibration models could be increasing rapidly if the proper preprocessing methods are applied (de Noord 1994). Sometimes it may not necessarily improve the predictive ability, but the parsimonious of the model. The author applied multiplicative signal correction (MSC) to the second derivative spectra of heavy oil products. The results demonstrated decreasing in model complexity and increasing in variance modeled. The model was expected to exhibit more robustness.

Ergon et al. studied on model choice and squared prediction errors (SPE) in PLS regression concerning residual and prediction consistency (Ergon, Halstensen et al. 2011).

One of the examples was to determine protein content in corn using NIR spectra with additional correction of prediction errors. Aimed at applying NIR spectroscopy to process monitoring systems, SPE caused by spectral process faults were added as **y** errors, which are persistent in industrial cases. The alternative non-orthogonalized PLS, bidiagonalization algorithms and re-formulation of the NIPALS algorithm were recommended instead of conventional PLS.

In agricultural industry, NIR is also widely applied on animal feed for industrial quality control purpose with on-line measurements. Research was conducted to study robust preprocessing methods and model selection for spectral data of animal feed samples (Verboven, Hubert et al. 2012). The authors proposed robust versions of the most well-known preprocessing methods such as standardization, first and second derivatives, MSC, SNV, detrending and smoothing techniques. A desirability index (DI) was created to evaluate the robust RMSEP, which is a one-number summary of a range of scores on different dimensions. Then, identify the robust model with the lowest RMSEP and the optimal number of latent variables. These parameters help to perform the model selection procedure in an automatic way and develop for routine use. It allows saving time and cost by simplifying the model development process in industry.

1.4.12 Economic analysis

NIR is an indirect technique that requires calibration based on the reference values from wet chemistry analysis, which is a considerable cost to users and calibration labs. The development of calibration on small datasets, maintenance and updating calibration only with new samples that bring in variations provides a way to reduce the cost and energy on calibration efforts instead of recalibration. For quality assurance purpose, the total cost in the NIR measurement system should be considered without sacrificing analytical accuracy in real application. In the grain industry, sampling and sample evaluation enables the feed manufacturers to make interference about the quality of incoming grain and decision of purchase. Statistical and economic analysis should be applied to a quality program to determine sampling and evaluation performance in NIR system.

References

Abdelkader, M. F., J. B. Cooper, et al. (2012). "Calibration transfer of partial least squares jet fuel property models using a segmented virtual standards slope-bias correction method." Chemometrics and intelligent laboratory systems **110**(1): 64-73.

Abe, H., C. Iyo, et al. (2000). "A study on the universality of a calibration with sample temperature compensation." Journal of Near Infrared Spectroscopy **8**(3): 209-213.

Abrahamsson, C., J. Johansson, et al. (2003). "Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets." Chemometrics and Intelligent Laboratory Systems **69**(1): 3-12.

Aminian, M. and F. Aminian (2000). "Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor." Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on **47**(2): 151-156.

Anderson, R. E., J. F. Hair, et al. (2006). Multivariate data analysis, Pearson.

Andrew, A. and T. Fearn (2004). "Transfer by orthogonal projection: making near-infrared calibrations robust to between-instrument variation." Chemometrics and Intelligent Laboratory Systems **72**(1): 51-56.

AOAC 990.03 (1995). Official Methods of Analysis. Washington, DC

Balabin, R. M., R. Z. Safieva, et al. (2008). "Wavelet neural network (WNN) approach for calibration model building based on gasoline near infrared (NIR) spectra." Chemometrics and intelligent laboratory systems **93**(1): 58-62.

Basheer, I. and M. Hajmeer (2000). "Artificial neural networks: fundamentals, computing, design, and application." Journal of microbiological methods **43**(1): 3-31.

Blanco, M., J. Coello, et al. (2001). "Orthogonal signal correction in near infrared calibration." Analytica Chimica Acta **434**(1): 125-132.

Bouveresse, E., C. Hartmann, et al. (1996). "Standardization of near-infrared spectrometric instruments." Analytical Chemistry **68**(6): 982-990.

Bouveresse, E., D. Massart, et al. (1994). "Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm effects of different standardisation samples." Analytica Chimica Acta **297**: 405-416.

Bouveresse, E. and D. L. Massart (1996). "Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration." Chemometrics and intelligent laboratory systems **32**(2): 201-213.

Brown, S. D. (2009). 3.08 - Transfer of Multivariate Calibration Models. Comprehensive Chemometrics. J. Kalivas, L. Buydens and P. van Espen. Oxford, Elsevier. **3:** 345-378.

Büning-Pfaue, H. (2003). "Analysis of water in food by near infrared spectroscopy." Food Chemistry **82**(1): 107-115.

Burns, D. A. and E. W. Ciurczak (2007). Handbook of near-infrared analysis, CRC.

Cai, W., Y. Li, et al. (2008). "A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra." Chemometrics and intelligent laboratory systems **90**(2): 188-194.

Capron, X., B. Walczak, et al. (2005). "Selection and weighting of samples in multivariate regression model updating." Chemometrics and intelligent laboratory systems **76**(2): 205-214.

Cassells, J. A., R. Reuss, et al. (2007). "Near infrared spectroscopic studies of changes in stored grain." Journal of near Infrared Spectroscopy **15**(3): 161-167.

Chalus, P., S. Walter, et al. (2007). "Combined wavelet transform–artificial neural network use in tablet active content determination by near-infrared spectroscopy." Analytica chimica acta **591**(2): 219-224.

Chau, F., Y. Z. Liang, et al. (2004). Chemometrics: from basics to wavelet transform, Wiley-Interscience.

Clark, R. D. (1997). "OptiSim: an extended dissimilarity selection method for finding diverse representative subsets." Journal of chemical information and computer sciences **37**(6): 1181-1188.

Dardenne, P. and R. Biston (1991). Standardization procedure and NIR instrument network. 3rd International Conference on Near Infrared Spectroscopy, Gembloux, Belgium, Agricultural Research Centre Publishing.

Daszykowski, M., K. Kaczmarek, et al. (2007). "Robust statistics in data analysis — A review: Basic concepts." Chemometrics and intelligent laboratory systems **85**(2): 203-219.

Daszykowski, M., B. Walczak, et al. (2002). "Representative subset selection." Analytica Chimica Acta **468**: 91-103.

Davies, A. M. C. and T. Fearn (2006). "Quantitative analysis via near infrared databases: comparison analysis using restructured near infrared and constituent data-deux (CARNAC-D)." Journal of Near Infrared Spectroscopy **14**(6): 403-411.

de Aguiar, P. F., B. Bourguignon, et al. (1995). "D-optimal designs." Chemometr. Intell. Lab. Syst. **30**(2): 199-210.

de Noord, O. E. (1994). "The Influence of Data Preprocessing on the Robustness and Parsimony of Multivariate Calibration Models." Chemometrics and intelligent laboratory systems **23**(1): 65-70.

Delwiche, S. R., R. O. Pierce, et al. (1998). "Protein content of wheat by near-infrared spectroscopy of whole grain: Collaborative study." Journal of Aoac International **81**(3): 587-603.

Despagne, F. and D. L. Massart (1998). "Neural networks in multivariate calibration." Analyst **123**(11): 157r-178r.

Despagne, F., B. Walczak, et al. (1998). "Transfer of calibrations of near-infrared spectra using neural networks." Appl Spectrosc **52**(5): 732-745.

Du, W., Z. P. Chen, et al. (2011). "Maintaining the predictive abilities of multivariate calibration models by spectral space transformation." Anal Chim Acta **690**(1): 64-70.

Duponchel, L., C. Ruckebusch, et al. (1999). "Standardisation of near infrared spectrometers using artificial neural networks." Journal of Near Infrared Spectroscopy **7**(3): 155-166.

Ergon, R., M. Halstensen, et al. (2011). "Model choice and squared prediction errors in PLS regression." Journal of Chemometrics **25**(6): 301-312.

Faber, N. K. M. (1999). "A closer look at the bias–variance trade off in multivariate calibration." Journal of Chemometrics **13**(2): 185-192.

Faber, N. K. M. (1999). "Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration." Chemometr. Intell. Lab. Syst. **49**(1): 79-89.

Fearn, T. (2000). "On orthogonal signal correction." Chemometrics and Intelligent Laboratory Systems **50**(1): 47-52.

Fearn, T. and A. M. C. Davies (2003). "A comparison of Fourier and wavelet transforms in the processing of near infrared spectroscopic data: Part 1. Data compression." Journal of Near Infrared Spectroscopy **11**(1): 3-15.

Ferre, J., S. D. Brown, et al. (2001). "Improved calculation of the net analyte signal in inverse multivariate calibration." Journal of Chemometrics **15**(6): 537-553.

Ferre, J. and F. X. Rius (1997). "Constructing D-optimal designs from a list of candidate samples." Trac-Trends in Analytical Chemistry **16**(2): 70-73.

Ferré, J. and F. X. Rius (1996). "Selection of the best calibration sample subset for multivariate regression." Analytical Chemistry **68**(9): 1565-1571.

Feudale, R. N., N. A. Woody, et al. (2002). "Transfer of multivariate calibrationmodels: a review." Chemom. Intell. Lab. Sys **64**: 181-192.

Flaten, G. R. and A. D. Walmsley (2003). "Using design of experiments to select optimum calibration model parameters." Analyst **128**(7): 935-943.

Flåten, G. R. and A. D. Walmsley (2004). "A design of experiment approach incorporating layered designs for choosing the right calibration model." Chemometrics and intelligent laboratory systems **73**(1): 55-66.

Foca, G., C. Ferrari, et al. (2011). "Minimisation of instrumental noise in the acquisition of FT-NIR spectra of bread wheat using experimental design and signal processing techniques." Anal Bioanal Chem **399**(6): 1965-1973.

Galvao, R. K., M. C. Araujo, et al. (2005). "A method for calibration and validation subset partitioning." Talanta **67**(4): 736-740.

Gemperline, P. (2006). Practical guide to chemometrics, CRC.

Guenard, R. D., C. M. Wehlburg, et al. (2007). "Importance of prediction outlier diagnostics in determining a successful inter-vendor multivariate calibration model transfer." Appl Spectrosc **61**(7): 747-754.

Guthrie, J., D. Reid, et al. (2005). "Assessment of internal quality attributes of mandarin fruit. 2. NIR calibration model robustness." Crop and Pasture Science **56**(4): 417-426.

Haaland, D. M. and D. K. Melgaard (2000). "New prediction-augmented classical least-squares (PACLS) methods: Application to unmodeled interferents." Applied Spectroscopy **54**(9): 1303-1312.

Haswell, S. J. (1992). Practical guide to chemometrics, CRC Press.

Hildrum, K. I. (1992). Near infra-red spectroscopy: bridging the gap between data analysis and NIR applications, Ellis Horwood Ltd.

Honorato, F. A., R. K. H. Galvão, et al. (2005). "Robust modeling for multivariate calibration transfer by the successive projections algorithm." Chemometrics and intelligent laboratory systems **76**(1): 65-72.

Höskuldsson, A. (2001). "Variable and subset selection in PLS regression." Chemometrics and intelligent laboratory systems **55**(1): 23-38.

Igne, B. and C. R. Hurburgh (2008). "Standardisation of near infrared spectrometers: evaluation of some common techniques for intra- and inter-brand calibration transfer." Journal of Near Infrared Spectroscopy **16**(6): 539-550.

Igne, B. and C. R. Hurburgh (2010). "Using the frequency components of near infrared spectra: optimising calibration and standardisation processes." Journal of Near Infrared Spectroscopy **18**(1): 39-47.

Igne, B., J. M. Roger, et al. (2009). "Improving the transfer of near infrared prediction models by orthogonal methods." Chemometrics and Intelligent Laboratory Systems **99**(1): 57-65.

Indahl, U. G. and T. Naes (1998). "Evaluation of alternative spectral feature extraction methods of textural images for multivariate modelling." Journal of Chemometrics **12**(4): 261-278.

Isaksson, T. and T. Næs (1990). "Selection of samples for calibration in near-infrared spectroscopy. Part II: Selection based on spectral measurements." Appl Spectrosc **44**(7): 1152-1158.

Janik, L. J., D. Cozzolino, et al. (2007). "The prediction of total anthocyanin concentration in red-grape homogenates using visible-near-infrared spectroscopy and artificial neural networks." Analytica Chimica Acta **594**(1): 107-118.

Jouan-Rimbaud, D., D. L. Massart, et al. (1997). "Characterisation of the representativity of selected sets of samples in multivariate calibration and pattern recognition." Anal Chim Acta **350**(1-2): 149-161.

Kalivas, J. H., N. Roberts, et al. (1989). "Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry." Analytical Chemistry **61**(18): 2024-2030.

Kawano, S., H. Abe, et al. (1995). "Development of a calibration equation with temperature compensation for determining the Brix value in intact peaches." Journal of Near Infrared Spectroscopy **3**: 211-218.

Kennard, R. W. and L. A. Stone (1969). "Computer aided design of experiments." Technometrics: 137-148.

KOSHOUBU, J., T. IWATA, et al. (2001). "Elimination of the Uninformative Calibration Sample Subset in the Modified UVE (Uninformative Variable Elimination)-PLS (Partial Least Squares) Method." Analytical sciences **17**(2): 319-322.

Li, H., Y. Liang, et al. (2009). "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration." Anal Chim Acta **648**(1): 77-84.

Lucasius, C. B., M. L. M. Beckers, et al. (1994). "Genetic algorithms in wavelength selection: a comparative study." Analytica Chimica Acta **286**(2): 135-153.

Maeda, H., Y. Ozaki, et al. (1995). "Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds." Journal of Near Infrared Spectroscopy **3**: 191-202.

Mallat, S. (2009). A wavelet tour of signal processing, Academic press.

Marini, F. (2009). 3.14-Neural Networks. Comprehensive Chemometrics. J. Kalivas, L. Buydens and P. van Espen. Oxford, Elisver. **3:** 477-505.

Martens, H. and S. Jensen (1982). Partial least squares regression: a new two-stage NIR calibration method. Proceedings of the Seventh World Cereal Bread Congress. Elsevier, Amsterdam, Netherlands.

Martens, H. and T. Naes (1992). Multivariate calibration, Wiley.

Mittermayr, C. R., H. W. Tan, et al. (2001). "Robust calibration with respect to background variation." Appl Spectrosc **55**(7): 827-833.

Næs, T. (1987). "The design of calibration in near infra red reflectance analysis by clustering." Journal of Chemometrics **1**(2): 121-134.

Naes, T., C. Irgens, et al. (1986). "Comparison of linear statistical methods for calibration of NIR instruments." Applied statistics **35**(2): 195-206.

Næs, T., T. Isaksson, et al. (2002). Standardisation of instruments. A user friendly guide to multivariate calibration and classification. T. Næs, T. Isaksson, T. Fearn and T. Davies. UK, NIR Publications**:** 207-219.

Næs, T. and H. Martens (1988). "Principal component regression in NIR analysis: viewpoints, background details and selection of components." Journal of chemometrics **2**(2): 155-167.

Norgaard, L., A. Saudland, et al. (2000). "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy." Appl Spectrosc **54**(3): 413-419.

Osborne, B. G. (2000). "Recent developments in NIR analysis of grains and grain products." Cereal Foods World **45**(1): 11-15.

Osborne, B. G. and T. Fearn (1983). "Collaborative Evaluation of near-Infrared Reflectance Analysis for the Determination of Protein, Moisture and Hardness in Wheat." Journal of the Science of Food and Agriculture **34**(9): 1011-1017.

Osborne, B. G. and T. Fearn (1983). "Collaborative Evaluation of Universal Calibrations for the Measurement of Protein and Moisture in Flour by near-Infrared Reflectance." Journal of Food Technology **18**(4): 453-460.

Osborne, B. G., Z. Kotwal, et al. (1999). "Optical matching of near infrared reflectance monochromator instruments for the analysis of ground and whole wheat." Journal of Near Infrared Spectroscopy **7**(3): 167-178.

Park, K. S., Y. H. Ko, et al. (2001). "Near-infrared spectral data transfer using independent standardization samples: a case study on the trans-alkylation process." Chemometrics and intelligent laboratory systems **55**(1-2): 53-65.

Peirs, A., N. Scheerlinck, et al. (2003). "Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents." Postharvest Biology and Technology **30**(3): 233-248.

Peng, J., S. Peng, et al. (2011). "Near-infrared calibration transfer based on spectral regression." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **78**(4): 1315-1320.

Pérez-Marín, D., A. Garrido-Varo, et al. (2005). "Implementation of LOCAL algorithm with near-infrared spectroscopy for compliance assurance in compound feedingstuffs." Appl Spectrosc **59**(1): 69-77.

Pierce, R., D. Funk, et al. (1996). Applying near infrared spectroscopy to the needs of US grain inspection.

Pieters, S., W. Saeys, et al. (2013). "Robust calibrations on reduced sample sets for API content prediction in tablets: Definition of a cost-effective NIR model development strategy." Analytica Chimica Acta **761**(0): 62-70.

Puchwein, G. (1988). "Selection of calibration samples for near-infrared spectrometry by factor analysis of spectra." Analytical Chemistry **60**(6): 569-573.

Puigdomènech, A., R. Tauler, et al. (1997). "Modelling near infrared instrument differences by chemometric methods: testing for near infrared forage analysis." Anal Chim Acta **355**(2–3): 181-193.

Roger, J. M., F. Chauchard, et al. (2003). "EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits." Chemometrics and Intelligent Laboratory Systems **66**(2): 191-204.

Roussel, S. A., C. L. Hardy, et al. (2001). "Detection of Roundup Ready (TM) soybeans by near-infrared spectroscopy." Appl Spectrosc **55**(10): 1425-1430.

Roussel, S. A., B. Igne, et al. (2011). "Noise robustness comparison for near infrared prediction models." Journal of Near Infrared Spectroscopy **19**(1): 23-36.

Sáiz-Abajo, M. J., B. H. Mevik, et al. (2005). "Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data." Analytica Chimica Acta **533**(2): 147-159.

Šašic, S. and Y. Ozaki (2001). "Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment." Analytical chemistry **73**(1): 64-71.

Setarehdan, S. K., J. J. Soraghan, et al. (2002). "Maintenance of a calibration model for near infrared spectrometry by a combined principal component analysis–partial least squares approach." Analytica Chimica Acta **452**: 35-45.

Shenk, J. S. (1991). "Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy." Crop science **31**(2): 469.

Shenk, J. S. (1991). "Population structuring of near infrared spectra and modified partial least squares regression." Crop science **31**(6): 1548.

Shenk, J. S. and M. O. Westerhaus (1985). "Accuracy of Nirs Instruments to Analyze Forage and Grain." Crop science **25**(6): 1120-1122.

Shenk, J. S. and M. O. Westerhaus (1989). Optical instrument calibration system. United States, 4866644.Sep 12,1989

Shenk, J. S., M. O. Westerhaus, et al. (1985). "Calibration Transfer between near-Infrared Reflectance Spectrophotometers." Crop science **25**(1): 159-161.

Shi, G. T., L. J. Han, et al. (2010). "Near infrared spectroscopy calibration transfer for quantitative analysis of fish meal mixed with soybean meal." Journal of Near Infrared Spectroscopy **18**(3): 217-223.

Siano, G. G. and H. C. Goicoechea (2007). "Representative subset selection and standardization techniques. A comparative study using NIR and a simulated fermentative process UV data." Chemometrics and intelligent laboratory systems **88**(2): 204-212.

Singh, C., J. Paliwal, et al. (2006). Near-infrared spectroscopy: applications in the grain industry. 2006 ASAE Annual Meeting, Paper.

Sjöblom, J., O. Svensson, et al. (1998). "An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra." Chemometrics and Intelligent Laboratory Systems **44**(1): 229-244.

Smith, B. M. and P. J. Gemperline (2000). "Wavelength selection and optimization of pattern recognition methods using the genetic algorithm." Analytica Chimica Acta **423**(2): 167-177.

Snee, R. D. (1977). "Validation of regression models: methods and examples." Technometrics: 415-428.

Sohn, M., F. E. Barton, II, et al. (2007). "Transfer of near-infrared calibration model for determining fiber content in flax: effects of transfer samples and standardization procedure." Appl Spectrosc **61**(4): 414-418.

Sun, D. W. (2009). Infrared spectroscopy for food quality analysis and control, Academic Press.

Swierenga, H., A. P. de Weijer, et al. (1999). "Strategy for constructing robust multivariate calibration models." Chemometrics and Intelligent Laboratory Systems **49**(1): 1-17.

Swierenga, H., W. Haanstra, et al. (1998). "Comparison of two different approaches toward model transferability in NIR spectroscopy." Appl Spectrosc **52**(1): 7-16.

Swierenga, H., F. Wülfert, et al. (2000). "Development of robust calibration models in near infra-red spectrometric applications." Analytica Chimica Acta **411**(1–2): 121-135.

Swingler, K. (1996). Applying neural networks: a practical guide. San Francisco, Morgan Kaufmann.

Tan, H. W. and S. D. Brown (2001). "Wavelet hybrid direct standardization of near-infrared multivariate calibrations." Journal of Chemometrics **15**(8): 647-663.

Tan, H. W. and S. D. Brown (2002). "Wavelet analysis applied to removing non-constant, varying spectroscopic background in multivariate calibration." Journal of Chemometrics **16**(5): 228-240.

Trygg, J. and S. Wold (2002). "Orthogonal projections to latent structures (O - PLS)." Journal of Chemometrics **16**(3): 119-128.

Verboven, S., M. Hubert, et al. (2012). "Robust preprocessing and model selection for spectral data." Journal of Chemometrics **26**(6): 282-289.

Walczak, B., E. Bouveresse, et al. (1997). "Standardization of near-infrared spectra in the wavelet domain." Chemometrics and intelligent laboratory systems **36**(1): 41-51.

Walczak, B. and D. L. Massart (1997). "Wavelet packet transform applied to a set of signals: A new approach to the best-basis selection." Chemometrics and intelligent laboratory systems **38**(1): 39-50.

Wang, B., G. Liu, et al. (2009). "Orthogonal projection to latent structures combined with artificial neural networks in non-destructive analysis of Ampicillin powder." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **71**(5): 1695-1700.

Wang, W. B., Y. X. Su, et al. (2002). "Standardization of Acousto-Optic Tunable Filter NIR spectrometric instrument." Optical Design and Testing **4927**: 56-63.

Wang, Y., D. J. Veltkamp, et al. (1991). "Multivariate instrument standardization." Analytical Chemistry **63**(23): 2750-2756.

Wang, Y. D. and B. R. Kowalski (1993). "Temperature compensating calibration transfer for nir filter instruments." Analytical Chemistry **65**: 1301-1303.

Wang, Y. D., M. J. Lysaght, et al. (1992). "Improvement of Multivariate Calibration through Instrument Standardization." Analytical Chemistry **64**(5): 562-564.

Watari, M. and Y. Ozaki (2006). "Practical calibration correction method for the maintenance of an on-line near-infrared monitoring system for molten polymers." Appl Spectrosc **60**(5): 529-538.

Westad, F. and H. Marten (2000). "Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression." Journal of Near Infrared Spectroscopy **8**(2): 117-124.

Williams, P. and K. Norris (1987). Near-infrared technology in the agricultural and food industries, American Association of Cereal Chemists, Inc.

Williams, P., K. Norris, et al. (1982). "Influence of temperature on estimation of protein and moisture in wheat by near-infrared reflectance." Cereal Chemistry **59**(6): 473-477.

Wold, S., H. Antti, et al. (1998). "Orthogonal signal correction of near-infrared spectra." Chemometrics and Intelligent Laboratory Systems **44**(1–2): 175-185.

Wold, S., M. Sjöström, et al. (2001). "PLS-regression: a basic tool of chemometrics." Chemometrics and intelligent laboratory systems **58**(2): 109-130.

Wulfert, F., W. T. Kok, et al. (2000). "Correction of temperature-induced spectral variation by continuous piecewise direct standardization." Analytical Chemistry **72**(7): 1639-1644.

Wulfert, F., W. T. Kok, et al. (2000). "Linear techniques to correct for temperature-induced spectral variation in multivariate calibration." Chemometrics and intelligent laboratory systems **51**(2): 189-200.

Zeaiter, M., J. M. Roger, et al. (2006). "Dynamic orthogonal projection. A new method to maintain the on-line robustness of multivariate calibrations. Application to NIR-based monitoring of wine fermentations." Chemometrics and Intelligent Laboratory Systems **80**(2): 227-235.

Zeaiter, M., J. M. Roger, et al. (2004). "Robustness of models developed by multivariate calibration. Part I: The assessment of robustness." TrAC Trends in Analytical Chemistry **23**(2): 157-170.

Zhang, L., G. W. Small, et al. (2002). "Calibration standardization algorithm for partial least-squares regression: application to the determination of physiological levels of glucose by near-infrared spectroscopy." Anal Chem **74**(16): 4097-4108.

Zhang, X., M. Chang, et al. (2010). "Influence and correction of temperature on optical measurement for fat and protein contents in a complex food model system." Infrared Physics & Technology **53**(3): 177-181.

CHAPTER 2. CALIBRATION SAMPLE SELECTION FROM A LARGE DATA POOL OF SOYBEAN IN NEAR-INFRARED SPECTROSCOPY

A paper to be submitted to Journal of Chemometrics

Nanning Cao[a], Charles R. Hurburgh[a]

[a] Iowa State University, Agricultural and Biosystems Engineering, Ames, Iowa 50010, USA.

Abstract

Near Infrared (NIR) calibration models for composition of biological materials are typically built from hundreds or thousands of samples over many years. To a point, a greater number of samples in the calibration provide more complete information. However, an extreme increase in the calibration set can lead to inadequate models. A large number of samples in Partial Least Squares (PLS)-based NIR models can reduce the prediction accuracy due to the over-inclusion of noise. Collecting large numbers of reference values by chemical methods is expensive. The development of a pre-selection routine for representative samples in a calibration is required to improve cost efficiency whiling maintaining robustness. Our goal for this paper was to identify samples contributing useful information to the calibration model, while eliminating redundant samples. Our case study is the calibration of soybean moisture, protein and oil, using a long term database (2001 crop through 2010 crop) containing more than 8,000 spectra plus references for moisture and 1,300 spectra plus reference for both protein and oil. We compared uniform random, Kennard-Stone algorithm and D-optimal method for selecting calibration samples for a PLS model. We made the assumption that non-linear modeling was not necessary in this case. It was identified that reliable NIR calibration can be obtained based on few representative samples ($n \approx 100$) by significantly reducing more than 98% of the original moisture data set and 90% of the protein and oil calibration set. Sample selection methods achieved the optimal subsets with about 400 spectra (7%, 32% and 40% of the original data for moisture, protein and oil, respectively), which were then used for calibration. All the samples from 2010 were used as the independent validation. Models built on the selected subsets provided equal or lower standard error of predictions (SEPs) of 0.14%, 0.53% and 0.54% for moisture, protein and oil, respectively when compared with the prediction results of models using the complete sets of calibration samples.

Keywords: sample selection; calibration optimization; near infrared spectroscopy

Introduction

The optimization of multivariate calibration models for near infrared spectroscopy (NIRS) is an iterative process searching for an optimal matrix with the most informative variables and samples for modeling with least additional noise. Variable selection studies have excluded irrelevant wavelength ranges of spectral data for partial least squares (PLS) regression. Methods that have improved the prediction performance of calibration models include the regression coefficients [1], interval-PLS [2], multi-objective genetic algorithm (GA) [3] and uninformative variable elimination (UVE) [4]. Few studies have reported on selection of calibration subsets from a larger data pool. The number of samples that could be used in calibration has increased steadily over time with reference of sample acquisition and data analysis. However, along with the development of the technique of instrumentation and the accumulation of the database of samples, the complexity of calibration has increased. Increasing spectral resolution increases the number of data points, which are not all directly relevant to the factor of interest. In the past decades, the meaning of "large data" has changed substantially in terms of both the number of observations (n) and variables (k). Large volumes of data (n>500, k>200) demonstrate inadequacies of multivariate calibration methods in both efficiency and interpretability [5]. When dealing with a large data set, it may be undesirable to include the whole data set in the calibration. Although more samples may provide more information that can reduce over fitting the model, redundant samples also increase the risk of adding more noise than information. When the number of samples is considered to be large (n>6k) compared to the number of variables, sample selection can help identify the most representative samples with diverse properties capable of creating a robust model. A smaller but more representative sample set would also increase the cost efficiency of the calibration model. Going forward, if the properties most responsible for making a sample "representative" can be identified in advance, costly reference chemistry can be saved. Ideally, only representative samples should be sent for reference chemical measurement.

Currently, three criteria are used for sample selection [6]. These include: (1) samples selected for calibration must contain all of the y and x variance expected in future prediction analysis. This means the range of variation in the concentration should be at least as wide as future prediction samples; (2) the samples must have uniform distribution of chemical components over their total range; and (3) a sufficient number of samples should be selected to statistically define the relationships between the spectral variables and the component concentrations or properties to be modeled. This ensures that the future analyses involve interpolation or extrapolation of the model. For some simple mixture products, it may be possible to prepare calibration samples to meet these criteria. For complex and natural products, it is difficult if not impossible to obtain a statistically ideal calibration set by controlling the compositions of samples in advance. Although these problems pose restrictions on the successful selection of a calibration sample set, they have received less attention than the topic of variable selection in the literature.

Previous studies have examined numerous feature selection methods for a representative calibration sample subset, but no generally accepted approach exists [7, 8]. Furthermore, these studies were based on small data sets of less than 1,000 original samples. Incorrectly, the rest of the samples after selection were used as validation sets. After selecting influential samples into calibration set, the remaining sample set represents a narrower range of variation and is no longer representative. Samples held from a larger pool automatically have variance correlated with the larger pool. Using such a validation set may result in optimistic prediction and optimization. Although larger databases have become available in the last several years, a common sample pool of sufficient size to validate the models has been lacking. Currently, there is no procedure to estimate the number of samples that should be included in calibration sets for the best possible future prediction.

The goal of this paper is to (1) choose the representative samples that cover the whole experimental region; (2) identify the optimal number of calibration samples and sample selection procedure to construct robust calibration models; (3) compare three sample selection algorithms in order to remain or improve the precision of calibration models built on data subsets from NIR transmittance measurements on whole soybean samples. In the present study, three approaches (uniform random selection, Kennard-Stone algorithm [9] and D-optimal method [10]), all based on multivariate statistics, were compared to select representative samples and to identify the optimal calibration set size for soybean sample spectra collected over the past ten years. With uniform designs, samples are selected uniformly distributed over the data space. The Kennard-Stone algorithm is commonly used in many applications. We investigated the effect of the number of samples in calibration sets on the prediction performance of calibration models. D-optimal criterion is based on the multivariate regression and experimental design. These methods were evaluated on their ability to optimize calibration models for soybean moisture, protein and oil contents in a near infrared transmittance instrument. A completely independent validation set of samples from the 2010 crop year was utilized to validate the calibration models.

Calibration Sample Selection

To develop a robust calibration model, calibration samples need to provide examples of all chemical components which are expected to be present in the future. The range of the variation in the concentrations of chemical components needs to exceed the range of variation expected for samples. Also, the concentration of chemical components of calibration samples should be uniformly distributed over their total range of variation [6].

Sample selection

For each calibration model, the original sample set was first sorted by their chemical reference values of each chemical component. Then, divide samples into subsets with increments of one percentage point of chemical reference values. Samples that had extremely low or high reference values were automatically kept in the calibration set. The three sample selection methods were applied to the remaining samples within each interval to generate different sizes of subsets. In each interval, we selected progressive numbers of samples. A total of 10 subsets were created for each component by one specific method. Table 1 describes the number of samples included in each interval for the three components. When the number of samples in the interval is smaller than the target number, include them all and no selection procedure need to be applied. Then, calibration sets with increasing sample numbers were obtained. The number of PCs was determined by cross validation. For each component, the same number of factors was applied to the series of models with increasing numbers of samples in each calibration set. This selection procedure enables a relatively uniform distribution of samples in terms of the content of chemical components. As the number of samples in each increment increased, the overall distribution approached that of the data pool.

Uniform random selection

Random selection is the simplest way to choose samples from a large population. However, this would lead to a normal distribution of reference values in the calibration set, more samples around the mean, and fewer extreme values as that of the original data pool. With uniform random selection, the total range of reference values was divided into several intervals. Within each interval, a certain amount of samples were randomly selected. When combined, the calibration set should have a relatively even distribution.

Kennard-Stone method

The Kennard-Stone algorithm (Kennard and Stone 1969) is designed to select samples sequentially and uniformly distributed over the object space by choosing samples that maximize the Euclidean distances between each other. Designed for surface response experimental plans, the distances are computed between the sample characteristics such as spectra. First, the two farthest samples are selected. Then, the sample farthest from the first two samples is selected as the third sample. Samples are then chosen one by one from the remaining subset, which is the farthest from all the previously selected until the desired size of the training set is reached. Assuming that k objects have already been selected (k<n, n is the number of samples), the (k+1)th object in the calibration set is chosen using the criterion

$$\max_{k<r\leq n}\left(\min\left(d^{2}{}_{1r,} d^{2}{}_{2r,\ldots} d^{2}{}_{kr}\right)\right) \tag{57}$$

where

$$d_{jr}=\left\|x_{j}-x_{r}\right\|=\sqrt{\sum_{j=1}^{k}\left(x_{j}-x_{r}\right)^{2}} \tag{58}$$

d_{jr} denotes the Euclidean distance from a candidate object r, not yet in the representative set, to the jth already selected object. This algorithm chooses the sample that presents the largest minimum distance with respect to any sample already selected at each successive iteration.

A modification of the Kennard-Stone algorithm has been proposed for dividing a data set into calibration and validation subsets [11]. This method defined the Euclidean distances of both spectral data and reference data. By dividing dx by their maximum values in the data set, equal importance was assigned to x and y spaces. A normalized **xy** distance $\mathbf{d}_{xy}$ was used instead of dx, which is used alone in Kennard-Stone algorithm.

However, samples selected by Kennard-Stone algorithm have a tendency to focus on the median ranges of concentration with a lower possibility of selecting samples with extreme concentrations [12]. Therefore, in this study, the Kennard-Stone algorithm was performed on each interval of samples as divided by their reference (y) values only.

D-optimal method

D-optimal designs are used for experimental design when the experimental region does not have a regular shape. Its principal objective is to maximize the determinant of the variance-covariance matrix |X'X|, where X is the training set matrix (p selected samples, m wavelengths) [10, 13]. This determinant is maximal when the selected samples span the space of the entire data set. The algorithm starts with a large data set and chooses samples iteratively which creates a maximum increase in the variance-covariance matrix $\left|\mathbf{X}'\mathbf{X}\right|$, and stops when chosen samples no longer increase the variance-covariance matrix $\left|\mathbf{X}'\mathbf{X}\right|$. To avoid getting a local optimal subset, the D-optimal algorithm must be run multiple times.

Materials and Methods

Materials

The original calibration set included approximately 9,000 whole soybean samples with references value for moisture, 1360 with protein and oil references. Samples were scanned at the Grain Quality Laboratory, Ames, IA, USA, during the crop years 2001-2009. Soybean samples from the crop year 2010 were used as validation set with 876 reference values for moisture, and 136 for protein and oil contents.

Spectra Acquisition

Soybean samples were scanned by one Bruins OmegAnalyzerG (serial number106110) (Bruins Instruments, Puchheim, Germany), a transmittance instrument with a spectra range from 730 to 1100 nm and an increment of 0.5 nm. The spectral range was reduced to from 850 to 1048 nm with a 2nm increment by a software option. A path length of 30mm was used. Samples were run at room temperature (22 ± 2 °C).

Reference chemical analyses

Moisture content was determined by the air oven method (AOCS Ac 2-41) [14] at Grain Quality Lab, Ames, Iowa State University. Protein content was determined by the combustion method (AOAC 990.03) [15] and oil content was determined by ether extract (AOCS Ac3-44) (AOCS 1998), both by Eurofins Scientific, Inc., Des Moines, Iowa, USA. Concentration values for protein and oil were converted to a 13% moisture basis. The reproducibility of the samples of these methods, as measured by blend replicates over the time period was 0.45 and 0.10 for protein and oil, respectively. Summary statistics for the calibration and validation sets are presented in Table 2.

Calibration procedure

Spectral pretreatment and outlier detection

Spectra were automatically mean centered when exported from the instrument. Second derivative was performed to remove constant and linear background by Savitzky-Golay algorithm (15-point window and third-order polynomial) including a smoothing step at the same time [16]. Standard normal variate (SNV) was applied following second derivative as a method for scatter correction to reduce the physical variability such as particle size between samples and adjust for baseline shift over the long period of data collection [16, 17]. The instrument had not been repaired on a significant way over the period. In order to detect outliers, PLS (leave-one-out cross validation) was carried out for each individual year to identify outliers and to check each year's chemistry. For years having few samples, samples were joined to the ones from the next year.

Leverage and residuals are the most important values for detecting outliers [18]. Leverage is a term to identify those observations that are far away from corresponding average predictor values. The basic tools for outlier detection in this model are based on the influence plot in Unscrambler 9.8 (Camo Software, Woodbridge, NJ) with leverage as x-axis and PCA residual as y-axis. The criteria for deleting outliers were: (1) Samples with PCA residuals higher than 2; (2) Samples with leverage greater than 3 times the average leverage. The average leverage (H) is calculated as:

$H = (1 + \text{number of principal components})/n$ (n is the number of samples) [19, 20]

Model development

PLS regression was used to develop all prediction models. In PLS, the original matrix is compressed into latent variables (LVs) that maximize the covariance between the reference values and all possible linear functions of the spectral data. PLS is a classical analytical technique to handle the multivariate nature of agricultural products and the highly collinear NIR spectroscopic data [5].

Programs

Calculations were performed with MATLAB R2011a (The MathWorks, Natick, MA) installed with the PLS_Toolbox v.6.2.1 (Eigenvector Research, Wenatchee, WA). All three sample selection methods including uniform random, Kennard-Stone and D-optimal were calculated in a custom-created Matlab program.

Validation procedure

An ideal validation set sample set should contain samples of all chemical components covers the range of variation in the concentration. Previous research indicated the impact of the variability of the next year's samples on the calibration process [21]. To maintain an independent and representative validation set, soybean samples from crop year 2010 were used as validation set. No outlier detection process was applied to the validation set. Thus, the validation set may comprise outlier samples and can be used to test the effectiveness of any outlier detection included in the calibration. The validation can also give an indication of the type and response if an outlier was inadvertently included in future results.

Calibration performance was be evaluated in terms of precision, accuracy, and model fit. The standard error of prediction (SEP) or standard deviation of differences and the relative predictive determinant (RPD) will be used to evaluate the precision [22]. The accuracy will be determined by the bias (average of differences). Bias is a good indicator of similarity between validation samples and the calibration set [23].

Uniformity of calibration set

The two-sided one-sample Kolmogorov-Smirnov test [24] was applied to determine whether the reference values of specific calibration sets differ significantly from a theoretical uniform distribution sets. Each calibration set was tested against the uniform distribution at level of $\alpha=0.05$. They were based on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution. The Kolmogorov-Smirnov test was performed in R (http://www.r-project.org/).

Results and Discussion

Description of initial models

Initial benchmark models were built using all the samples in the data pool. The predictive abilities of these models are summarized in the last row of Table 3. The results of these models for three contents are precise and accurate. This may owe to the coverage of the large calibration sets that lead to a calibration model with high variability and robust to variations in the new crop year. However, when the original dataset is too large, it is not computational efficient to include them all into calibration set. From the aspect of model updating and maintenance, it is not practical to keep adding all the new samples to the data pool. Thereby, an objective selection of the optimum combination of samples that contain the greatest variation could facilitate the development of a robust calibration model. Statistics of these initial models were used to evaluate statistics of subsets and selection methods.

Calibration sample selection

After examining the structure of the moisture content of the validation set, we found there were too many unusual samples in the original pool with extremely high moisture contents. Most of these samples came from crop year 2009, which was an unusually wet year. Therefore, only samples with a more typical moisture content ranging from 4% to 18% were used as the original pool for sample selection. The same number (9) PCs was used in all models for moisture. Theoretically, as more samples were included, additional PLS factors could be revealed due to new sources of variance. However, this was not the case for all three components derived from the prediction results. Models with different sample sizes for each component used the same number of PLS factors.

Comparing SEP and RPD values for the models of moisture content (Table 3(a).), models with calibration sets selected by uniform random method and D-optimal had slightly different prediction abilities. With the least number of samples, D-optimal gives the lowest SEP value. Except KS, uniform random and D-optimal showed lower SEP value with increasing number of samples in the calibration set. The SEPs approached the benchmark results and then outperformed with around four hundreds samples included. Calibration set selected by KS showed similar trend, but without satisfying results. This means the number of samples is independent on selection method. Unlike KS and D-optimal, the subset selected by uniform random is not unique because of the random procedure. It is required to run this algorithm several times to obtain stable results. In terms of computational efficiency, D-optimal is recommended to choose the calibration subset when the data set it selects from is large.

For protein content, staring with the original set of 1349 samples, uniform random method selected a series of calibration sets which lead to models with equal or better predictive performances than the original sample set. A number of 12 PCs was utilized for all the PLS models. SEP and RPD values for the series of calibration models of protein content were shown in Table 3(b). Models built on sample sets selected by Kennard-Stone method and D-optimal method demonstrated similar prediction properties, close to the performance of the benchmark.

The SEP and RPD values of oil content in Table 3(c), showed that no models built on the selected subsets outperformed the benchmark. SEPs of models built on subsets selected by uniform random and D-optimal methods decreased with an increasing of sample numbers and approached the optimal SEP obtained by the original set. Nine PCs were used to build all PLS models.

Representative properties of selected samples

By selecting spectra which display the maximum variance, these subsets represent samples in the domain of the original pool. For moisture, protein and oil, the first two PCs explained 95.15%, 93.75% and 93.99% of the total variance, respectively. Thereby, the first two PCs scores of selected and eliminated samples were depicted in Figure 1 for visualization. These samples were selected by D-optimal, uniform random and Kennard-Stone for moisture, protein and oil, respectively. PCA score plots of all the sample spectra showed no clear clusters according to harvest years. Selected samples spread out in the spectral domain in the first two PCs. Figure 1 gives the smallest subset size examples for each of the three methods. This demonstrated that the original domain of all the spectra was well covered by selected subsets even with the least number of samples included.

Figure 2, 3 and 4 give an example of the histograms of reference values for the three components of all the available samples (a) and the selected smallest subsets (b). Figures (2(a), 3(a) and 4(a)) of the original data pool all showed bell shapes, which suggested normal distributions. This finding is reasonable due to the nature of agriculture products. Samples tended to have content values accumulate near the mean. Fewer samples naturally showed up with extreme contents.

Figure 2(b) provides the histogram of soybean moisture content for 140 samples selected by D-optimal algorithm. This subset displayed a flat distribution on the chemical measurements and representative on the **y** space (reference values). For moisture content, all the distribution of subsets in the calibration models were not significantly different from the uniform distribution at level of $\alpha=0.05$. This may due to the large number of available samples in the original pool.

The distribution of protein content of 113 samples selected by uniform random method shown in Figure 3(b) is flat. This subset were not significantly different from the uniform distribution at level of $\alpha=0.05$. For oil content, Kennard-Stone algorithm selected a subset of 129 samples with a uniform distribution (at level of $\alpha=0.05$, Figure 4(b)) on reference values. When the number of samples in the extreme intervals (low or high reference values) becomes smaller than the target number, increasing sample size only increased samples close to the mean. Then, the distribution of the rest subsets lost their uniformity.

The results of Kolmogorov-Smirnov test indicated that the smallest subsets selected by these three approaches obtained uniform distributions of the chemical measurements. They are considered to be representative in terms of both X (spectra) and Y (reference values) domain. By eliminating most samples with reference values close to the mean, more robust models could be achieved on calibration subsets by avoiding subsequent analysis to regress toward the mean.

Identification of sample size in calibration set

Patterns of SEPs observed in Figure 5 suggested that further extension of the calibration subsets give less improvement on SEP. It became stable when an optimal amount of samples were included. With additional similar samples, unrelated information reduced the robustness of calibration models and reduced the accuracy of prediction in the validation set.

With only around 100 samples included in calibration sets, comparable results were obtained for all three components. This is only 2%, 10% and 17% of the total number of samples with moisture, protein and oil contents, respectively. The optimal number of samples was about 400 to 500, with which moisture and protein models outperformed the benchmark. Beyond this point, no significant improvement of SEP was observed. From a practical point of view, if the goal is to choose a subset of samples that could be reliably used for model construction without any compromise in its predictive ability, it is recommended to include the optimal number of samples. Nevertheless, calibration on about 100 representative samples is the most cost-efficient approach.

Comparison of calibration models

Since all the calibration models were tested on the same validation set, errors from the lab measurements for the validation set contribute to all the sets of prediction errors [25]. In terms of SEP, there were slight differences among the prediction performance based on the series of calibration models built on subsets selected by the three sample methods. Adequate calibration models were built on subsets reduced from 8462 to 140 samples for moisture, from 1349 to 113 samples for protein, and from 1339 to 129 samples for oil. For moisture and protein contents, models built on subsets with optimal number of samples improved the prediction performance significantly ($\alpha=0.05$). For oil content, models with a SEP of 0.56 was not statistically different from the benchmark of 0.53 ($\alpha=0.05$).

When comparing results of the three sample selection approaches, uniform random and D-optimal are recommended. Uniform random method has an advantage due to its random procedure, which enables more flexibility, while D-optimal is more computational efficient. Kennard-Stone algorithm did significant worse than the other two approaches for moisture content. When there was too many samples in each interval as for moisture samples, samples selected by KS may not be representative enough to the whole region. It selected new sample based on its distance to the selected ones, without considering the rest samples as a whole. It is clear that consideration of both X and Y information will obtain a more reasonable result of sample selection for a multivariate calibration.

Model accuracy

Model accuracy (bias) is an important factor to evaluate calibration models. The series of biases obtained for each factor illustrated error from the lab measurements. Based on long term quality control data from the Grain Quality Lab at Iowa State University, the bias of the reference measurements for soybean samples from the chemistry lab is 0.45% and 0.10% lower than the average values of several replicates for protein and oil, respectively. This part of error comes from the performance of the chemistry lab, which is consistent with the bias obtained in the calibration models. NIR models built on subsets demonstrated their abilities to calibrate time (crop year) as a variable with improved robustness and less samples.

Conclusions

Representative sample subsets were selected to spread over both the range of reference and spectral data domain. Calibration models based on these subsets were able to obtain similar or even better performance on prediction. The least number of samples required for a robust calibration is 100.

The optimal numbers of calibration samples were identified. For moisture content, 490 samples selected by uniform random constructed the best model with a SEP of 0.15%, comparing to 0.20% and 0.17% by KS and D-optimal, respectively. For protein content, uniform random selected 429 samples to obtain SEP of 0.53%, while the optimal results in KS and D-optimal are 0.55% (n=571) and 0.56% (n=503), respectively. As for oil content with 388 samples, SEPs for uniform random, KS and D-optimal were 0.54%, 0.55% and

0.55%. Numbers of samples in calibration sets were significantly reduced to 6% of the original 8,462 for moisture and 30% of the original 1,300 for protein and oil contents, which may save substantial expense for reference analysis.

Uniform random, Kennard-Stone and D-optimal methods showed similar results in selecting representative calibration sets. No model significantly outperformed the others in terms of predictive ability from a practical aspect. SEPs became stable after reaching an optimum in the prediction of every constituent. This illustrates the feasibility of constructing model on small calibration sets without compromising on robustness. The uniform random method gave slightly smaller SEPs with the same number of samples in calibration set.

Overall, the importance of uniformity in the calibration set was confirmed. The selection of representative samples helps avoid similar samples and redundant calculations, to achieve robust calibration models. There is no universal optimal number of samples for calibration set. The initial sample set drawn from the calibration set must be large and representative.

References

1. Westad, F. and Martens, H., Variable selection in near infrared spectrscopy based on significance testing in partial least squeares regression. J. of Near Infrared Spectrosc, 2000. 8(2): p. 117-124.

2. Norgaard, L., et al., Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc, 2000. 54(3): p. 413-419.

3. Shinzawa, H., et al., Multi-objective genetic algorithm-based sample selection for partial least squares model building with applications to near-infrared spectroscopic data. Appl Spectrosc, 2006. 60(6): p. 631-640.

4. Cai, W., Li, Y., and Shao, X., A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. Chemometr. Intell. Lab. Syst., 2008. 90(2): p. 188-194.

5. Kettaneh, N., Berglund, A., and Wold, S., PCA and PLS with very large data sets. Comput Stat Data Anal, 2005. 48(1): p. 69-85.

6. ASTM E1655-05, Standard Practices for Infrared Multivariate Quantitative Analysis, 2005: 100 Barr Harbor Dr.,West Conshohocken, PA 19428.

7. Isaksson, T. and Næs, T., Selection of samples for calibration in near-infrared spectroscopy. Part II: Selection based on spectral measurements. Appl Spectrosc, 1990. 44(7): p. 1152-1158.

8. Ferré, J. and Rius, F.X., Selection of the best calibration sample subset for multivariate regression. Analytical Chemistry, 1996. 68(9): p. 1565-1571.

9. Kennard, R.W. and Stone, L.A., Computer aided design of experiments. Technometrics, 1969: p. 137-148.

10. de Aguiar, P.F., et al., D-optimal designs. Chemometr. Intell. Lab. Syst., 1995. 30(2): p. 199-210.

11. Galvao, R.K., et al., A method for calibration and validation subset partitioning. Talanta, 2005. 67(4): p. 736-40.

12. Chen, D., Cai, W., and Shao, X., Representative subset selection in modified iterative predictor weighting (mIPW) — PLS models for parsimonious multivariate calibration. Chemometr. Intell. Lab. Syst., 2007. 87(2): p. 312-318.

13. Ferre, J. and Rius, F.X., Constructing D-optimal designs from a list of candidate samples. Trac-Trends in Analytical Chemistry, 1997. 16(2): p. 70-73.

14. AOCS Ac2-41, Ac3-44, Official methods and recommended practices of the American oil Chemists' Society, 1998, AOCS Press: Champaign.

15. Sweeney, R.A., Generic combustion method for determination of crude protein in feeds: collaborative study. Journal-Association of Official Analytical Chemists, 1989. 72(5): p. 770.

16. Barnes, R.J.D., M. S. Lister Susan J., Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. Appl Spectrosc, 1989. 43(5): p. 772-777.

17. Rinnan, Å., Berg, F.v.d., and Engelsen, S.B., Review of the most common pre-processing techniques for near-infrared spectra. Trends in Analytical Chemistry, 2009. 28(10): p. 1201-1222.

18. Daszykowski, M., et al., Robust statistics in data analysis — A review: Basic concepts. Chemometrics and intelligent laboratory systems, 2007. 85(2): p. 203-219.

19. Faber, N.K.M., Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration. Chemometr. Intell. Lab. Syst., 1999. 49(1): p. 79-89.

20. Faber, N.K.M., A closer look at the bias–variance trade off in multivariate calibration. Journal of Chemometrics, 1999. 13(2): p. 185-192.

21. Igne, B., et al., Influence of yearly variability of agricultural products on calibration process: a triticale example. Cereal Chemistry, 2007. 84(6): p. 576-581.

22. Igne, B. and Hurburgh, C.R., Local chemometrics for samples and variables: optimizing calibration and standardization processes. Journal of Chemometrics, 2010. 24(1-2): p. 75-86.

23. Kalivas, J.H., et al., Calibration Maintenance and Transfer Using Tikhonov Regularization Approaches. Appl Spectrosc, 2009. 63(7): p. 800-809.

24. Darling, D.A., The Kolmogorov-Smirnov, Cramer-von Mises Tests. The Annals of Mathematical Statistics, 1957. 28(4): p. 823-838.

25. Fearn, T., Comparing standard deviations. NIR news, 1996. 7(5): p. 5-6.

Table 1. Numbers of samples included in each interval of reference values

Subsets	Moisture	Protein	Oil
1	5	5	5
2	10	10	10
3	15	15	20
4	20	20	30
5	25	25	40
6	30	30	50
7	35	35	60
8	40	40	70
9	45	50	80
10	50	60	90

Table 2. Summary statistics of the calibration and validation sets

Constitutes	Model	N	Mean (%)	Range (%)	Standard deviation (% pts)
Moisture	Calibration [a]	8462	9.73	4.09-18.00	3.11
	Validation [b]	876	8.05	4.01-17.63	2.60
Protein (13% moisture basis)	Calibration [a]	1349	36.37	24.72-46.89	3.63
	Validation [b]	138	35.73	28.81-45.67	4.02
Oil (13% moisture basis)	Calibration [a]	1339	18.30	11.85-24.64	1.97
	Validation [b]	136	18.38	12.88-22.06	1.98

[a] 2001-2009 crop years

[b] 2010 crop year only

Table 3. Validation results of soybean moisture (a), protein (b) and oil (c) calibration models by sample set siz

(a)

N[a]	Uniform Random			Kennard-Stone method			D-optimal		
	SEP[b] (% pts)	Bias[c] (% pts)	RPD[d]	SEP (% pts)	Bias (% pts)	RPD	SEP (% pts)	Bias (% pts)	RPD
140	0.23	0.18	11.14	0.35	0.19	7.46	0.17	0.14	15.00
210	0.21	0.19	12.42	0.29	0.17	9.10	0.22	0.14	11.58
280	0.20	0.18	13.20	0.29	0.21	8.87	0.21	0.14	12.64
350	0.20	0.17	13.14	0.25	0.20	10.25	0.17	0.16	15.14
420	0.16	0.18	16.41	0.23	0.19	11.53	0.17	0.16	15.06
490	0.15	0.19	16.90	0.20	0.18	12.68	0.17	0.15	15.18
560	0.16	0.18	16.66	0.19	0.15	13.46	0.16	0.13	16.55
630	0.14	0.17	18.14	0.20	0.16	13.30	0.16	0.14	15.89
700	0.16	0.15	16.68	0.19	0.17	13.60	0.16	0.15	16.71
770	0.16	0.15	16.31	0.19	0.17	13.71	0.16	0.16	16.34
8462(All)	0.17	0.21	15.68	0.17	0.21	15.68	0.17	0.21	15.68

[a] Number of samples in calibration set

[b] Standard error of prediction

[c] SEP and Bias are expressed in % pt

[d] Relative predictive determinant

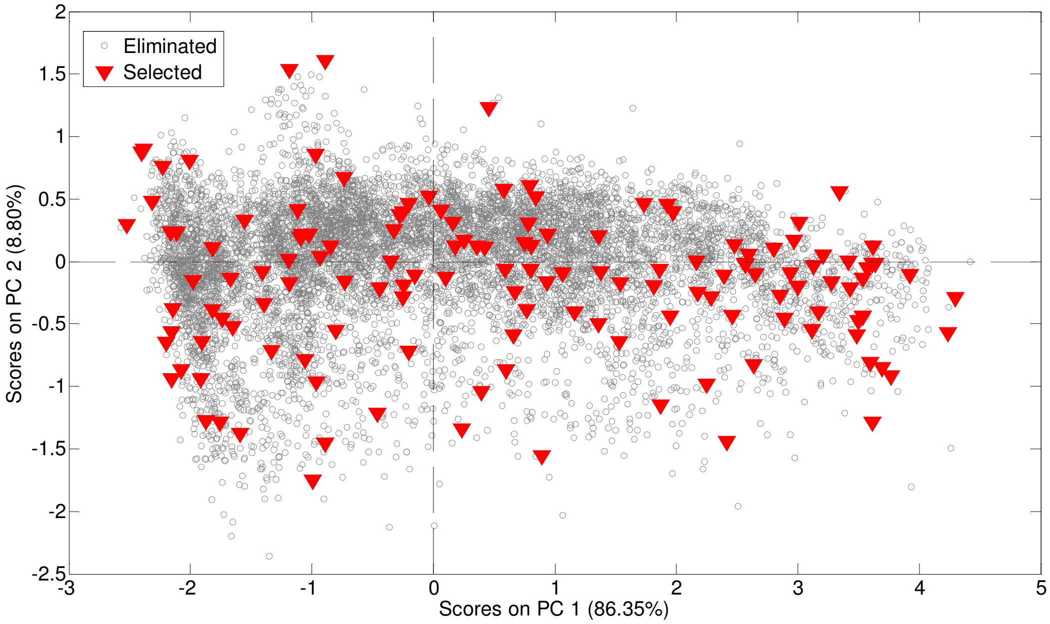
(b)

N	Uniform Random			Kennard-Stone method			D-optimal		
	SEP (% pts)	Bias (% pts)	RPD	SEP (% pts)	Bias (% pts)	RPD	SEP (% pts)	Bias (% pts)	RPD
113	0.54	0.21	7.38	0.63	0.08	6.37	0.60	0.19	6.63
203	0.55	0.31	7.28	0.63	0.21	6.38	0.60	0.19	6.64
279	0.57	0.30	7.06	0.61	0.28	6.56	0.61	0.30	6.54
353	0.58	0.34	6.91	0.59	0.29	6.73	0.57	0.32	7.05
429	0.53	0.35	7.53	0.58	0.32	6.91	0.58	0.31	6.85
503	0.59	0.37	6.71	0.57	0.34	7.05	0.56	0.34	7.07
571	0.57	0.35	6.98	0.55	0.36	7.21	0.58	0.33	6.91
624	0.54	0.31	7.43	0.55	0.35	7.28	0.56	0.33	7.08
719	0.55	0.34	7.20	0.56	0.35	7.13	0.58	0.37	6.90
799	0.53	0.34	7.52	0.56	0.37	7.14	0.58	0.39	6.94
1349(All)	0.56	0.42	7.13	0.56	0.42	7.13	0.56	0.42	7.13

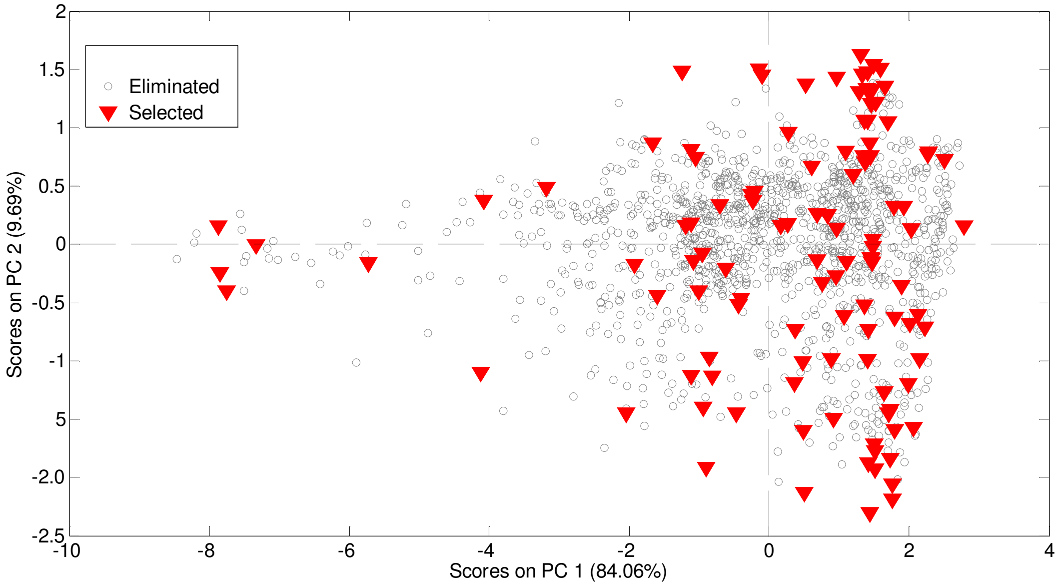
(c)

N	Uniform Random			Kennard-Stone method			D-optimal		
	SEP (% pts)	Bias (% pts)	RPD	SEP (% pts)	Bias (% pts)	RPD	SEP (% pts)	Bias (% pts)	RPD
129	0.58	0.14	3.41	0.56	0.05	3.55	0.58	0.14	3.53
228	0.55	0.02	3.60	0.59	0.09	3.40	0.56	0.07	3.57
308	0.55	0.04	3.60	0.57	0.05	3.49	0.56	0.06	3.57
388	0.54	0.05	3.70	0.55	0.05	3.61	0.55	0.07	3.62
468	0.55	0.04	3.63	0.55	0.07	3.62	0.55	0.06	3.65
542	0.53	0.04	3.72	0.55	0.06	3.61	0.55	0.06	3.63
604	0.53	0.03	3.74	0.55	0.06	3.62	0.55	0.06	3.65
664	0.54	0.03	3.70	0.55	0.06	3.64	0.55	0.06	3.66
724	0.54	0.02	3.70	0.54	0.04	3.68	0.54	0.04	3.69
776	0.53	0.00	3.76	0.54	0.05	3.67	0.55	0.05	3.66
1339(All)	0.53	-0.01	3.75	0.53	-0.01	3.75	0.53	-0.01	3.75

(a)


Eliminated
Selected
Scores on PC 2 (8.80%)
Scores on PC 1 (86.35%)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
-2
-1
0
1
2
3
4
5

(b)


Eliminated
Selected
Scores on PC 2 (9.69%)
Scores on PC 1 (84.06%)
2
1.5
1
0.5
0
-0.5
-1
5
-2.0
-2.5
-10
-8
-6
-4
-2
0
2
4

(c)

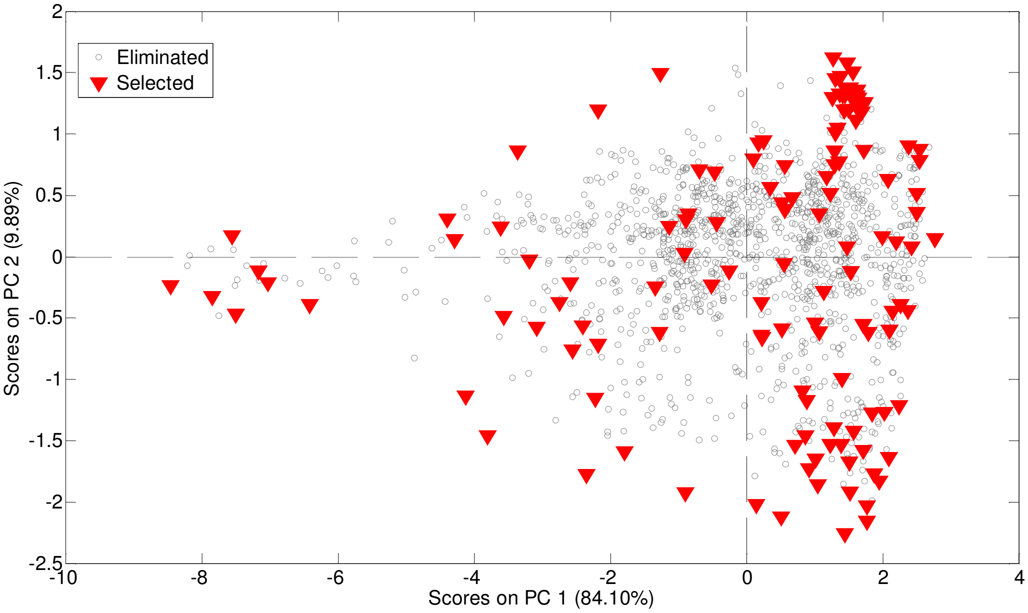
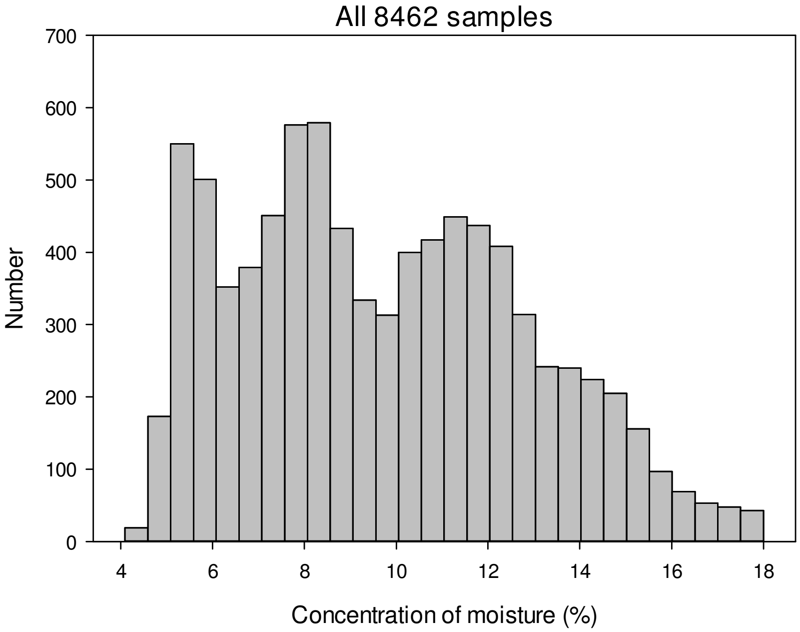


Figure 1. PCA score plot showing the space covered by minimum sample models (a) 140 samples using D-optimal algorithm for soybean moisture; (b) 113 samples using uniform random method for soybean protein; (c) 129 samples using Kennard-Stone algorithm for soybean oil

(a)



(b)

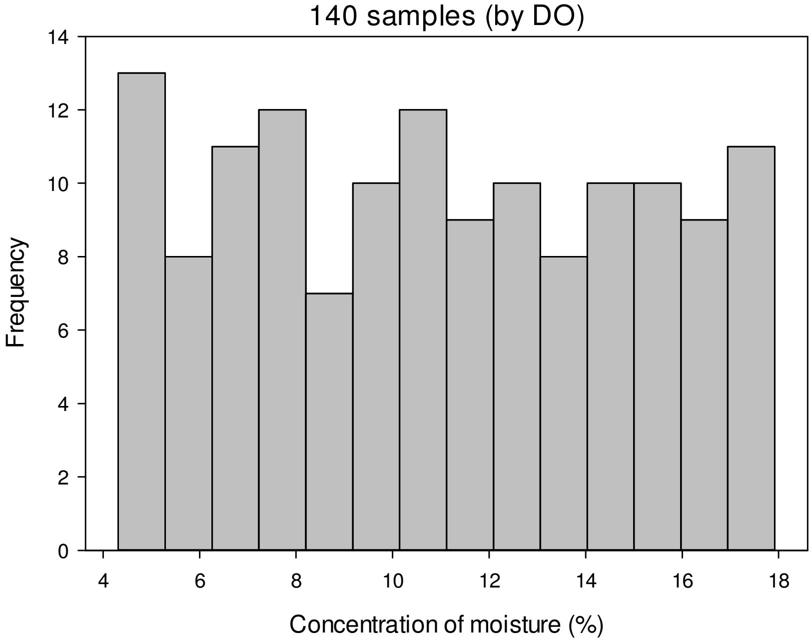
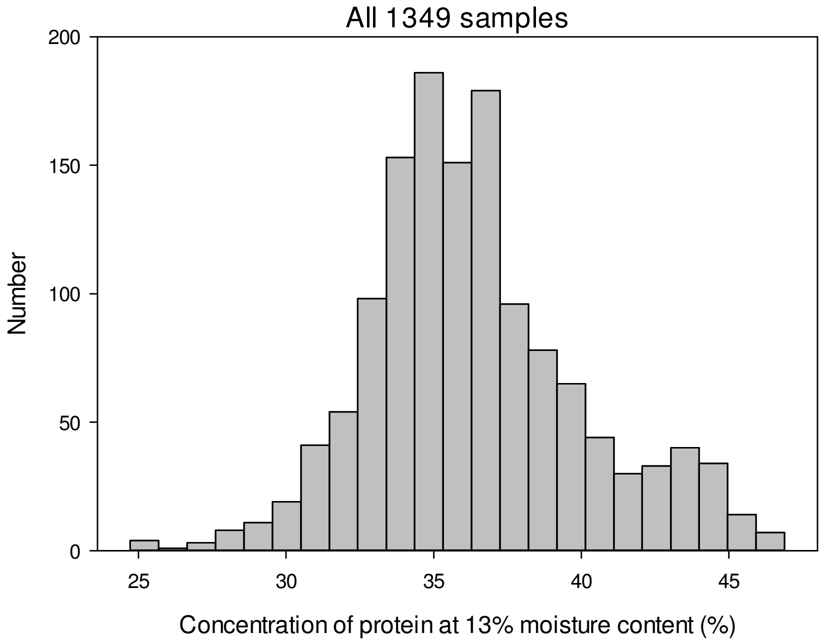


Figure 2. Distribution of all samples (n=8462) (a) and selected samples using D-optimal algorithm for chemical measurements (n=140) (b) for soybean moisture content

(a)



(b)

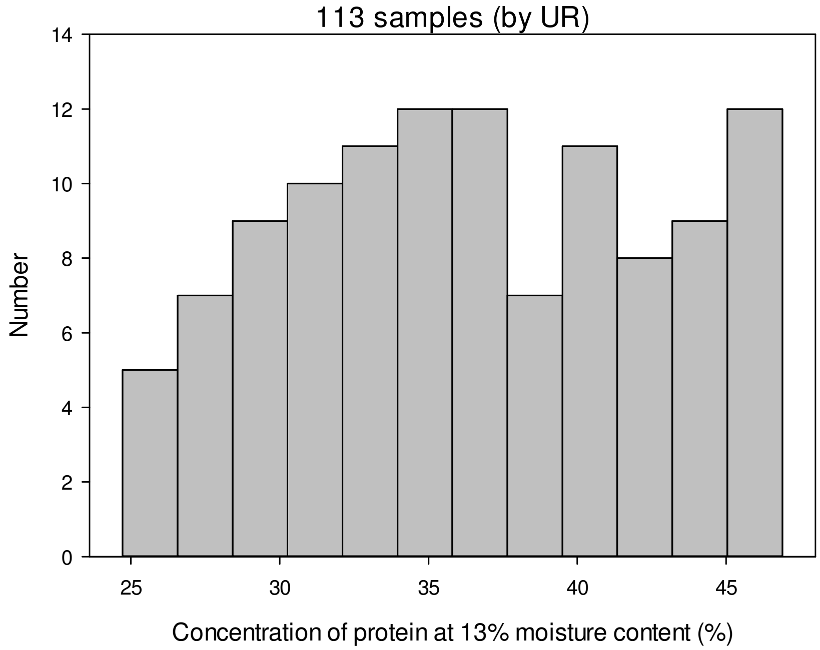
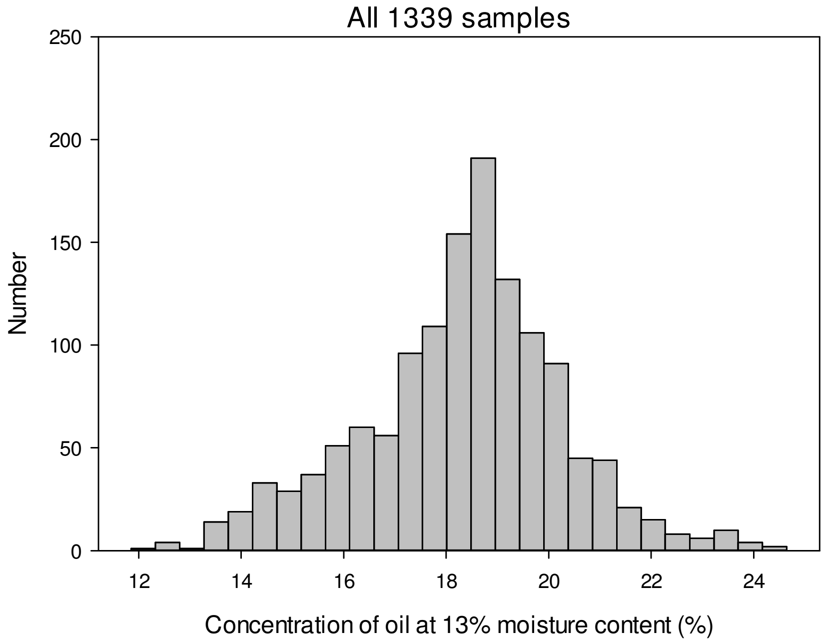


Figure 3. Distribution of all samples (n=1349) (a) and selected samples using uniform random method for chemical measurements (n=113) (b) for soybean protein content

(a)



(b)

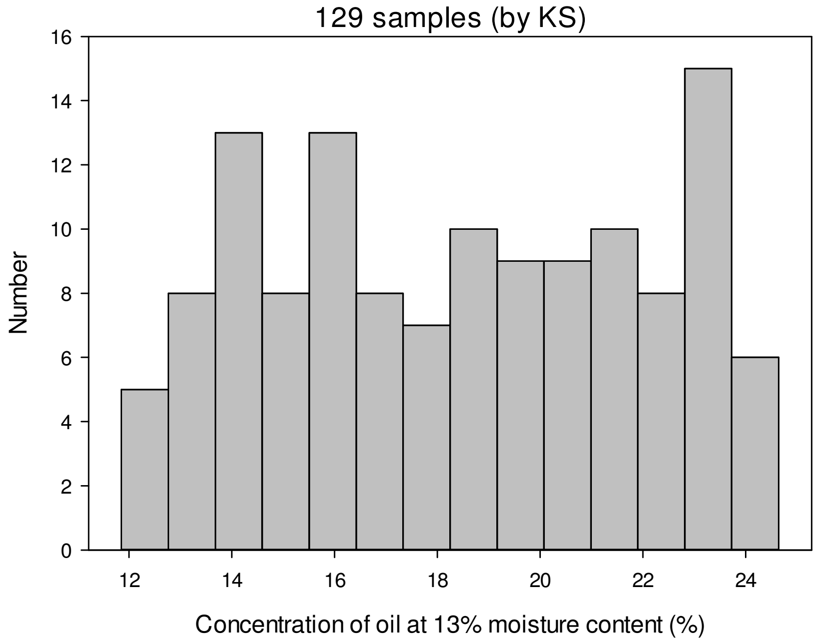
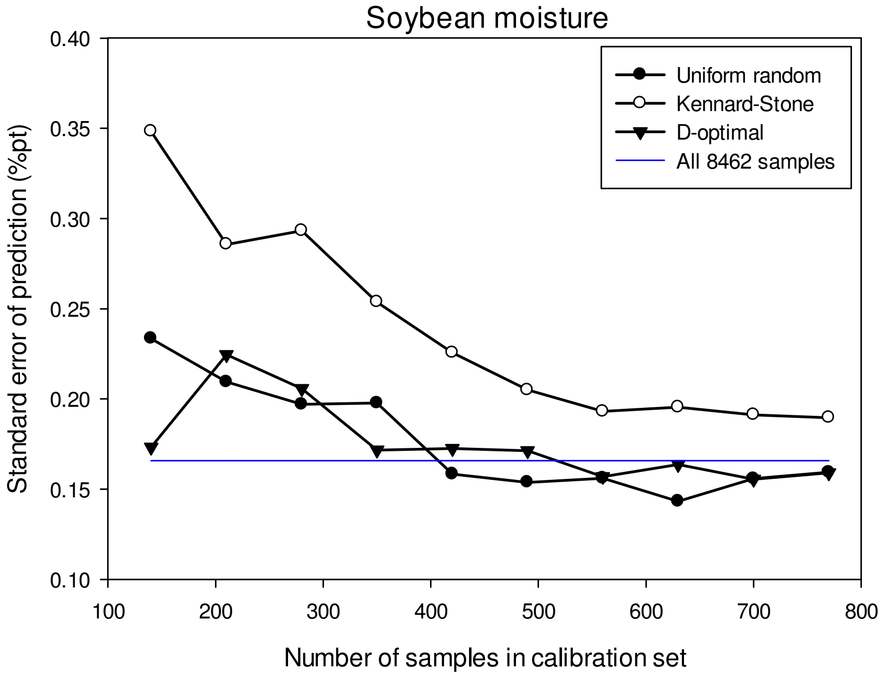
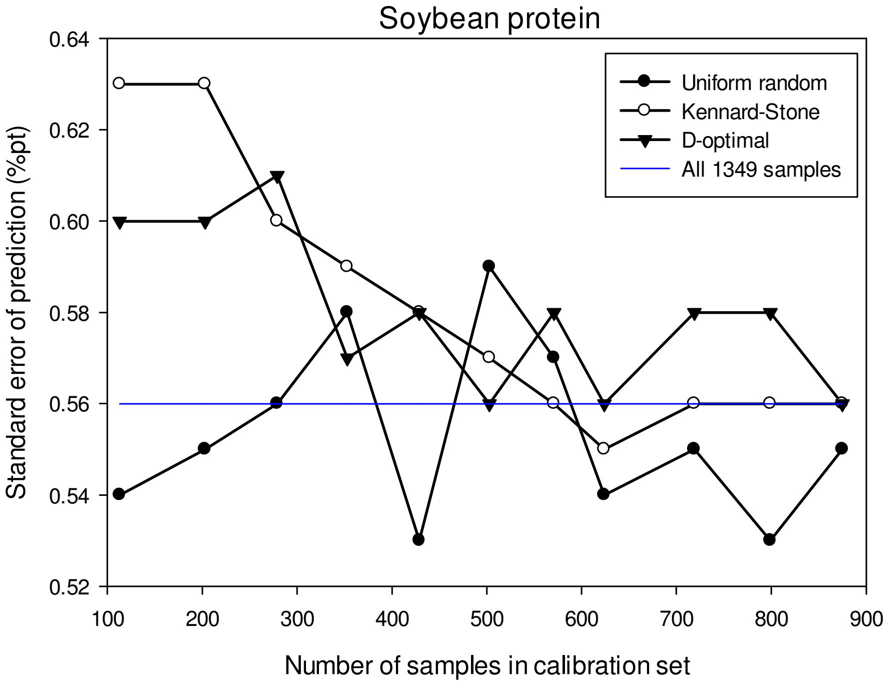


Figure 4. Distribution of all samples (n=1339) (a) and selected samples using Kennard-Stone algorithm for chemical measurements (n=129) (b) for soybean oil content

(a)


Soybean moisture
Standard error of prediction (%pt)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
100
200
300
400
500
600
700
800
Number of samples in calibration set
Uniform random
Kennard-Stone
D-optimal
All 8462 samples

(b)


Soybean protein
Standard error of prediction (%pt)
0.64
0.62
0.60
0.58
0.56
0.54
0.52
100
200
300
400
500
600
700
800
900
Number of samples in calibration set
Uniform random
Kennard-Stone
D-optimal
All 1349 samples

(c)

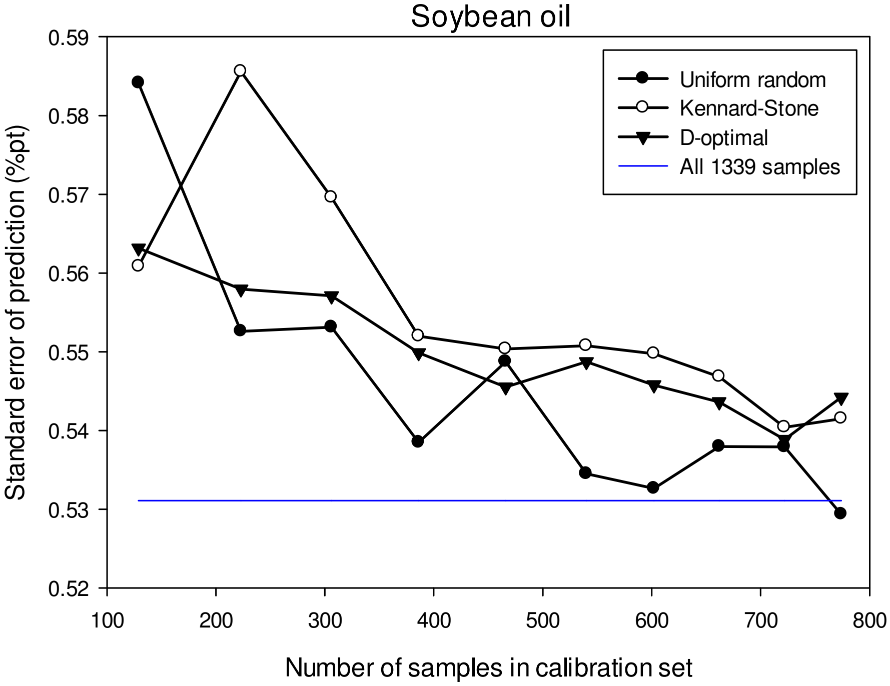


Figure 5. SEPs of soybean moisture (a), protein (b) and oil (c) contents by different calibration sample set sizes.

CHAPTER 3. OPTIMIZATION ON NIR NETWORKS: EFFECT OF CALIBRATION SETS ON INSTRUMENT STANDARDIZATION

A paper to be submitted to Chemometrics and Intelligent Laboratory Systems

Nanning Cao[a], Charles R. Hurburgh[a]

[a] Iowa State University, Agricultural and Biosystems Engineering, Ames, Iowa 50010, USA.

Abstract

In this study, the feasibility of using calibration subsets for instrument standardization was demonstrated. Whole soybean samples from crop year 2001 to 2011 were used as the original pool. Calibration samples were selected by D-optimal method to obtain representative subsets. Three levels of calibration subsets were created for comparison.

Calibration models were transferred from one instrument to another in a network of four transmittance instruments provided by two vendors. We compared three standardization methods for calibration transfer (no standardization, robust model and slope and bias correction). Soybean protein and oil were used for analysis. The prediction results showed it is possible to obtain comparable transferability performance with well-selected small data subsets (about 10% and 35% of the original numbers of samples) as using the entire dataset. The optimal calibration subset ($n \approx 400$) had similar or better performance in terms of calibration transfer compared with benchmark model. A simple slope and bias correction was recommended for standardization between similar instruments. This method provided the lowest SEPs of 0.48% and 0.34% for protein and oil, respectively on Bruins, 0.55% and

0.35%, respectively on Infratec. Representative calibration subsets ($n \approx 120$ and 400) reduced the complexity of the transferability of calibration model. Simplified calibration maintenance and model update procedure save cost of reference analyses and spectral measurements on multiple instruments, both of which are user limitations in practical application.

Keywords: Calibration transfer; Sample subsets; Calibration optimization

Introduction

In recent years, calibration transfer among instruments has become indispensable for the application of near infrared (NIR) spectroscopy calibration models. Calibration transfer enables NIR calibration models to be applied on multiple instruments at different locations working in union on common calibration equations. It would save calibration work if the database of agricultural products could be shared and provide quality specifications efficiently. Multivariate calibration models are intended to be used for an extended period of time. For agricultural products (grains, meals and flours), samples from the next crop year always add new variations to the calibration pool. Variations come from climate, variety, locations and other external factors. These factors add uncertainty to prediction process. To incorporate new variations into calibration models, two common approaches are available: create new calibration models or update existing calibration models by including samples with new variations to the original database. Compared to recalibration, model updating minimizes calibration effort, improves robustness and variability. Effective calibration transfer enables database sharing and is improved by samples collected across crop years, locations and varieties. Long-term maintenance of near infrared (NIR) calibration models keep pace with technology advancement and new sources of sample variations is a costly process. Using effective and appropriate calibration transfer methods to solve this long-standing problem is an opportunity for both researchers and industry workers searching for a competitive advantage.

Various published works [1-9] on calibration transfer focus on the methods of calibration transfer among copies of instruments. This is also called instrument standardization. The functions of standardization methods are to estimate the spectral differences among different instruments on varying measurement situations, therefore to correct the estimated differences with valid standardization parameters. Based on different types of instruments and standardization problems in a particular network, it is of great importance to apply appropriate standardization methods. For very similar instruments, standardizing the spectral response is mathematically more complex than standardizing the calibration models but provides better results as it accommodates slight spectral differences that could be corrected via simple calculations. When the differences among instruments are small, post-regression slop and bias correction (SBC) [1] could obtain reliable results of prediction. SBC is a simple and practical standardization method, which only requires a univariate correction. It corrects the predicted values based on the univariate linear model, yielding standardized predictions. For multivariate methods, commonly employed techniques are direct standardization (DS) and piecewise direct standardization (PDS) [4, 5, 10, 11]. In PDS, a moving window of neighboring wavelengths is used instead of the entire spectral range to correct differences between spectra (X-space).

Another approach to correct spectral differences is using signal preprocessing in order to improve robustness of resulting model. Preprocessing techniques such as multiplicative signal correction (MSC) [12], standardized normal variate (SNV) [13] can be used to remove systematic noise like baseline shift and scattering effects caused by different particle sizes. Orthogonal signal correction (OSC) [14] and its modified versions were developed to avoid information loss and retain information regarding the factor of interest (y values).

To compute valid standardization parameters of any form, standardization samples need to be well-chosen, stable and representative of both the calibration and prediction samples. They are scanned on multiple instruments and used to compute the standardization parameters. The selection of standardization samples is associated with standardization methods applied and different practical reasons such as sample properties. Standardization samples could be a subset of calibration set, prediction set or generic standards of a similar nature [15]. Sample selection methods such as Kennard and Stone algorithm [16] and D-optimal [17] provide a good and representative subset for the PDS procedure.

However, little research has been conducted to examine the effect of calibration set on instrument standardization aiming at a robust calibration. The robustness criterion of multivariate models is “the stability of its predictive capacity against perturbations centered on standard conditions” [18]. This means the calibration models should incorporate all possible variables of both calibration and prediction steps including external differences. For instrument recalibration, an alternative way to standardization is to construct a new calibration model by using only a small subset of the original training set. The most informative objects from the original pool were selected as calibration set. Sample selection plays an important role in calibration transfer for the purpose of model updating. Using a small subset of samples reduces the cost involved in the repetition of the analyses at the secondary instrument. The goal of calibration maintenance is to add in samples represent new variations and improve the robustness. To achieve the goal, the transfer sample set needs to be representative of the spectra and reference relationships in the data that must be captured by the multivariate calibration model. Especially in a large-scale spectral database, proper spectral management and database sharing could provide rapid analysis for complex mixtures, even as complex as agricultural products [19].

The objective of this study was to determine the relationship between the number of samples in a calibration set and subsequent calibration transfer performance. The optimal calibration set should model the parameters of interest comprehensively and accommodate instrumental variations with the fewest possible number of samples included. Protein and oil contents in soybeans were the test cases in a network using four NIR transmittance analyzers supplied by two vendors.

Materials and Methods

Materials

The original calibration set included approximately 1,100 whole soybean samples with reference values of protein and oil from one chemistry laboratory. Samples were scanned at Grain Quality Laboratory, Ames, IA, USA, during crop years 2001-2011. Soybean samples from crop year 2011 ($n \approx 150$) were used as the validation set.

Instrumentation

Soybean samples were scanned by Bruins OmegAnalyzerGs (serial numbers: 106110 and 106118, respectively) (Bruins Instruments, Puchheim, Germany) and two Foss Infratec Grain Analyzers (1229 and 1241, serial numbers 553075 and 12410350 respectively) (FOSS North America, Eden Prairie, MN, USA). They are transmittance instruments with effective spectra range from 850 to 1048 nm in 2nm increments. A path length of 30mm was used for both models. Samples were run at room temperature (22 ± 2 °C).

Calibration subsets

Calibration set includes samples are used to build a calibration model on the “master” instrument. Three levels of subsets were selected by D-optimal method [17] to reduce redundancy. D-optimal designs are used for experimental design when the experimental region does not have a regular shape. Its principal objective is to maximize the determinant of the variance-covariance matrix $|X'X|$, where X is the training set matrix (p selected samples, m wavelengths) [17, 20]. This determinant is maximal when the selected samples span the space of the entire data set. The algorithm starts with a large data set and chooses samples iteratively which creates a maximum increase in the variance-covariance matrix $|X'X|$ and stops when chosen samples no longer increase the variance-covariance matrix $|X'X|$. With this principle, this algorithm selects the subset of sample spectra that generate the largest space. For a two-dimensional space, it tends to select extreme samples on the edge of the data space. Different sample selection methods were compared in the previous study.

Model validation

An ideal validation set sample set should contain samples of all chemical components covering the range of variation in the concentration. Previous research indicated the impact of the variability of the next year’s samples on the calibration process [21]. To maintain an independent and representative validation set, soybean samples from crop year 2011 were used as validation set. No outlier detection process was applied to the validation set. Thus, the validation set may comprise outlier samples and can be used to test the effectiveness of any outlier detection included in the calibration. Samples from the prediction set are measured in the secondary instrument, and the spectra are transferred to be equivalent to the master instrument with the standardization parameters computed with the standardization set.

Calibration performance was evaluated in terms of precision, accuracy, and model fit. The standard error of prediction (SEP) or standard deviation of differences and the relative predictive determinant (RPD) were used to evaluate the precision [22]. The accuracy will be determined by the bias (average of differences). Bias is a good indicator of similarity between validation samples and the calibration set [23].

Calibration transfer set

The calibration transfer samples are required to cover the variability of the data in a representative manner. There are usually two ways to obtain calibration transfer set: (1) select the best subset samples from the set to be predicted; (2) use samples not coming from the prediction set, but of similar nature. In this study, 20 whole soybean samples selected from crop years 1996 - 2010 by the Grain Quality Laboratory, Ames, IA, USA were used as standardization set. These samples have multiple (75) reference measurements for each constituent and were not included in the calibration set. This standardization set is updated annually and has proven to be a good predictor over many years. They are used for post-regression slope and bias correction of instrument standardization, which is the only method the current instrument software allows. The other methods in this study were applied to the spectra downloaded to Matlab (The MathWorks, Natick, MA). Samples from the standardization set are measured on the two instruments to compute the standardization parameters.

Reference chemical analyses

Protein content was determined by the combustion method (AOAC 990.03) [24] and oil content was determined by ether extract (AOCS Ac3-44) [25], both by Eurofins Scientific, Inc., Des Moines, Iowa, USA. Concentration values for protein and oil were converted to a 13% moisture basis. Summary statistics for the calibration and validation sets are presented in Table 1.

Combined Preprocessing techniques

As a preprocessing technique, standardized normal variate (SNV) can improve scan repeatability as well as instrument similarity. It normalizes each spectrum to reduce scattering effects due to packing heterogeneity or path-length variations [26]. It is also known to improve the inter-instrument transferability. We considered the combination of SNV followed by detrending (DT). Detrending spectra account for the variation in baseline shift and linearity of spectra by using a second-degree polynomial to correct each spectrum [13].

Model construction and optimization

PLS regression was used to develop all prediction models. In PLS, the original matrix is compressed into latent variables (LVs) that maximize the covariance between the reference values and possible linear functions of the spectral data. PLS has been a classical analytical technique to handle the multivariate natures of the agriculture products and the highly collinear NIR spectroscopic data.

Standardization procedures

No standardization

In this approach, no standardization method was applied to the calibration model developed on the master instrument. Spectra scanned on the secondary instrument were directly transferred for prediction. The results this approach were compared with those of the original model and of the other standardization methods.

Robust models

To build a global robust model, the calibration should include all the external variations as well as the parameters of interest. In this case, the instrumental variations were calibrated into the model as a variable by combining measurements on both instruments.

This corporates all relevant sources of variation in the calibration design in order to develop a more or less universal calibration model [27].When spectral variation caused by factors different from the parameter to be predicted such as external environmental factors are present in calibration data, a common approach is to include this variation in the calibration model. For this purpose, the calibration sample spectra measured under standard conditions; the spectra of a smaller set measured under changed conditions are combined into one dataset global calibration model is constructed. With this principle, spectra collected on both master and secondary instruments were combined to incorporate instrumental differences.

Slope and bias correction (SBC)

This standardization method is a simple correction on predicted y-values for the standardization set with the calibration model. In this post regression method, it assumes a linear relationship between the prediction values for spectra measured on the secondary instrument and the prediction values that the obtained by calibration built on the master instrument. The properties of the standardization samples (y-values) were firstly predicted by the original model developed on the master instrument.

$$y_{master} = \mathbf{X}_{master} b \tag{59}$$

$$y_{secondary} = \mathbf{X}_{secondary} b \tag{60}$$

A linear regression equation is obtained by plotting predicted values collected in the calibration step against those collected in the prediction step using either ordinary least squares or orthogonal least squares. Then the predicted values for the new samples are corrected for the bias (intercept) and slope of the regression line:

$$y_{corrected} = bias + slope\, \mathrm{y}_{secondary} \tag{61}$$

This standardization approach is most often applied between instruments having the same dispersion device, which applies to our case of two inter-brand instruments. When more complex instrumental differences exist, other standardization methods need to be used. Complexity of calibration transfer process increases user cost very rapidly.

Software

Calculations were performed with MATLAB R2011a (The MathWorks, Natick, MA) with the PLS_Toolbox v.6.2.1 (Eigenvector Research, Wenatchee, WA). D-optimal sample selection method was computed in a custom-created Matlab program.

Experimental procedure

The four instruments in this network share the same original data pool of soybean samples from crop year 2001 to 2010. Each instrument was calibrated to its original calibration set with three different levels of numbers of samples. Then these calibration models were validated on the validation set. The robust models were developed by combining measurements on both instruments within brand. Samples in the standardization set were used to transfer the instrumental variances, while the robustness of transfer models were assessed by the independent validation set from crop year 2011.

Results and Discussion

This section shows comparisons among three standardization methods and three levels of calibration samples. These comparisons are made based on SEP and RPD values obtained by a series of calibration models. Tables 2 and 3 display the prediction results on the same validation set using calibrations built on four instruments of protein and oil contents, respectively. According to the prediction performance on each instrument of its own, OmegaAnalyzerG 106110 and Infratec Grain Analyzer 1241(0350) were assigned as master instruments.

Spectral responses

The averaged raw and preprocessed spectra of all available samples for OmegaAnalyzerG network did not show visible spectral differences, as shown in Fig. 1(a). Baseline shift (Fig. 2(a)) between the averaged spectra of two instruments in Infratec network was examined. After preprocessing, the averaged spectra of the same brand overlapped heavily in Fig. 1(b) and Fig. 2(b). Second derivative and SNV effectively eliminated the effects of baseline shift, scaling and scattering in Infratec instruments.

Calibration subsets

Based on an earlier study, three levels of calibration sets with increasing numbers of samples were used for comparison. For this particular study, the D-optimal method was applied for the selection of calibration subsets due to its ability choosing representative samples and computational effectiveness. Selected sample subsets possessed a wide range of reference values and uniform distribution. They were also well spread out in PCA score plot of spectral analysis that illustrated the spectral differences.

The first level contained the fewest samples (n≈120) selected from the original pool that demonstrated similar performance to the benchmark model (the entire dataset). Although for some calibrations compromises need to be made for lower RPD, the accuracy is acceptable for practical quantitative applications. The use of standardization methods always improved the results for calibrations built on the least number of samples as compared to the direct application of the calibrations built on the master instruments. The number of 100 calibration samples was the minimum in the master calibration necessary to cover the variability of the data in a representative manner.

The second level referred to the optimal numbers of samples that outperformed the benchmark. About 400 samples are included in this level. For OmegAnalyzerGs, the original models on the master instrument, both protein and oil models reached comparable results as the benchmark model. For standardization, the protein content at this level obtained better RPD than the benchmark model with SBC. Calibration for neither protein nor oil models on Infratec grain analyzers outperformed the benchmark models. However, after standardization, the differences among different numbers of calibration sets were minimized. Especially for protein content, RPDs were significantly improved compared to the original models on the secondary instruments themselves.

The third level included all available samples. This was the benchmark model. It contains about 1,100 samples for each constituent. Calibration models built on this level remained the best results in most cases. This was more obvious without standardization. However, after standardization, these SEPs were not statistically different from the ones obtained on the optimal number of samples. This means samples in the second level were representative enough to convey information of spectral and instrumental differences. More samples in the calibration models would not improve the performance of its transferability and prediction ability. Moreover, large dataset leads to less efficiency on calculation. This also applied to robust models when included spectra measured on both master and secondary instruments, which doubled the number of samples.

Standardization procedures

Figures 3 and 4 showed the RPD values for comparison of three standardization methods as well as the original calibrations on each instruments on their own for protein and oil content, respectively. In Figure 3, OmegAnalyzerGs showed better transfer ability than Infratec grain analyzers with all three standardization methods in terms of protein content. Comparable results were obtained by the two brands for oil content.

No standardization

Without applying any standardization method, calibrations developed on the two master instruments were directly used to predict spectra collected on the secondary instruments. Due to the similarity of the master and secondary instruments, there was no obvious wavelength drift. This approach received acceptable or even better results than the prediction performance of the calibration models built on the secondary instruments with some calibration subsets. Moreover, samples were collected under very close environmental situations. Thus, with a more robust calibration model built on the master instrument, it is possible to achieve better prediction. It is worth noting that calibration model for oil content in Bruins OmegAnalyzerG was not sensitive to the numbers of calibration samples. Small calibration sets worked as well as large ones.

The primary condition when samples measured on the master instruments was similar to the operating circumstances of the secondary instruments. Moreover, the prediction samples were measured in a similar instrumental and environmental state.

Robust models

A global model was created by combining spectra collected on both instruments of the same brand at all three levels of calibration subsets. For protein content, robust models improved the transfer ability of OmegAnalyzerGs, while not significantly on Infratec grain analyzers. For oil content, both brands obtained slightly increasing but non-significant RPD values. This indicates that with the same original pool in this network, there is no need to scan same samples on two similar instruments with the purpose of standardization.

Post-regression slope and bias correction (SBC)

The prediction values on the secondary instrument were corrected to match with the master instrument by using the 20 standardization samples. The linear relationship of the 20 standardization samples was plotted in Figures. Then, predicted values on secondary units were corrected by the slope and bias. SEPs were obtained between the corrected values and reference values. This approach is suitable for transfers between instruments using the same dispersion device.

The key consideration is that the RPD values of the calibrations built on sample subsets increased significantly after slope and bias correction for protein content on both brands. For oil content, calibration subsets achieved comparable transferability with the original pool on OmegAnalyzerGs. These results illustrated the potential of the transferability of the two levels of calibration subsets. Without obvious instrumental difference between the master and secondary unites, SBC is a simple and effective standardization method in practice. Its weaknesses are the need to retest the standard samples in each instrument copy if the calibration is updated, and the need for complete representativeness in the set.

Discussion

Overall, all the standardization methods applied in this study provided adequate calibration transfer from the master instrument to the secondary instrument, compared with its own calibration. For different numbers of calibration sets, the first level with about 100 samples provided the minimum necessary variability to cover the original pool. With sufficient samples, around 400 representative samples were able to obtain comparable results to the benchmark models. In terms of spectral difference, SNV and DT were able to remove baseline shift and scattering effect of the raw spectra and reduce the spectral differences between instruments.

For standardization methods, SBC improved the results of calibration model with the least number of calibration samples. For the optimal and original datasets, SBC was slightly better than no standardization and robust models. The advantage of slope and bias correction is it only requires a number of calibration transfer samples with known reference values to be scanned on the secondary instruments. This could save time and energy on scanning future samples in the prediction set.

Calibration maintenance and model update is an important procedure. The results demonstrated the possibility of accomplishing this task by updating calibrations on the master instruments with a representative calibration subset which could be transferred to secondary instruments.

Conclusions

This study examined the effect of numbers of calibration samples on calibration transfer (standardization). The prediction results after standardization demonstrated the feasibility of using subset samples for calibration transfer between similar instruments. The use of calibration transfer brought the prediction errors down to values comparable to those obtained with complete recalibration of the secondary instrument. The lowest SEPs of 0.48% and 0.34% for protein and oil, respectively on Bruins, 0.55% and 0.35%, respectively on Infratec were obtained by slope and bias correction.

This study illustrates the use of calibration subsets via three different standardization methods. Different levels of representative sample subsets allow the reduction in complexity of the transferability of calibration model. With fewer samples ($n \approx 120$ and 400) included in calibration, comparable transferable performances were achieved on both protein and oil contents. The optimal calibration subsets ($n \approx 400$) showed better prediction values (RPD>6.0 and 5.5 for protein and oil, respectively) than the benchmark calibration in all cases.

Even though all methods in certain circumstances provided precision improvements in standardization, slope and bias correction is a simple and effective method to improve the transferable performance of calibration model based on calibration subsets ($n \approx 120$ and 400). In this network, we recommend using simple slope and bias correction standardization method to transfer calibrations when the instrumental difference is simple. This method has the advantage over robust model of not requiring samples to be measured on both master and secondary instruments, which saves time and cost substantially and is more practical for routine use.

References

[1] Bouveresse, E., et al., Standardization of near-infrared spectrometric instruments. Analytical Chemistry, 1996. 68(6): p. 982-990.

[2] Shenk, J.S. and M.O. Westerhaus, Optical instrument calibration system, 1989: United States.

[3] Dardenne, P. and R. Biston. Standardization procedure and NIR instrument network. in 3rd International Conference on Near Infrared Spectroscopy. 1991. Gembloux, Belgium: Agricultural Research Centre Publishing.

[4] Wang, Y., D.J. Veltkamp, and B.R. Kowalski, Multivariate instrument standardization. Analytical Chemistry, 1991. 63(23): p. 2750-2756.

[5] Wang, Y.D., M.J. Lysaght, and B.R. Kowalski, Improvement of Multivariate Calibration through Instrument Standardization. Analytical Chemistry, 1992. 64(5): p. 562-564.

[6] Puigdomènech, A., et al., Modelling near infrared instrument differences by chemometric methods: testing for near infrared forage analysis. Anal Chim Acta, 1997. 355(2–3): p. 181-193.

[7] Despagne, F. and D.L. Massart, Neural networks in multivariate calibration. Analyst, 1998. 123(11): p. 157r-178r.

[8] Walczak, B., E. Bouveresse, and D.L. Massart, Standardization of near-infrared spectra in the wavelet domain. Chemometrics and intelligent laboratory systems, 1997. 36(1): p. 41-51.

[9] Fearn, T., Standardisation and calibration transfer for near infrared instruments: a review. J. Near Infrared Spectrosc, 2001. 9(4): p. 229-244.

[10] Wang, Y. and B.R. Kowalski, Calibration transfer and measurement stability of near-infrared spectrometers. Appl Spectrosc, 1992. 46(5): p. 764-771.

[11] Bouveresse, E. and D.L. Massart, Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. Chemometrics and intelligent laboratory systems, 1996. 32(2): p. 201-213.

[12] de Noord, O.E., The Influence of Data Preprocessing on the Robustness and Parsimony of Multivariate Calibration Models. Chemometrics and intelligent laboratory systems, 1994. 23(1): p. 65-70.

[13] Verboven, S., M. Hubert, and P. Goos, Robust preprocessing and model selection for spectral data. Journal of Chemometrics, 2012. 26(6): p. 282-289.

[14] Wold, S., et al., Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent Laboratory Systems, 1998. 44(1–2): p. 175-185.

[15] Bouveresse, E., D. Massart, and P. Dardenne, Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm effects of different standardisation samples. Analytica Chimica Acta, 1994. 297: p. 405-416.

[16] Kennard, R.W. and L.A. Stone, Computer aided design of experiments. Technometrics, 1969: p. 137-148.

[17] de Aguiar, P.F., et al., D-optimal designs. Chemometr. Intell. Lab. Syst., 1995. 30(2): p. 199-210.

[18] Zeaiter, M., et al., Robustness of models developed by multivariate calibration. Part I: The assessment of robustness. TrAC Trends in Analytical Chemistry, 2004. 23(2): p. 157-170.

[19] Zhou, W., Y. Ying, and L. Xie, Spectral Database Systems: A Review. Applied Spectroscopy Reviews, 2012. 47(8): p. 654-670.

[20] Ferre, J. and F.X. Rius, Constructing D-optimal designs from a list of candidate samples. Trac-Trends in Analytical Chemistry, 1997. 16(2): p. 70-73.

[21] Igne, B., et al., Influence of yearly variability of agricultural products on calibration process: a triticale example. Cereal Chemistry, 2007. 84(6): p. 576-581.

[22] Igne, B. and C.R. Hurburgh, Local chemometrics for samples and variables: optimizing calibration and standardization processes. Journal of Chemometrics, 2010. 24(1-2): p. 75-86.

[23] Kalivas, J.H., et al., Calibration Maintenance and Transfer Using Tikhonov Regularization Approaches. Appl Spectrosc, 2009. 63(7): p. 800-809.

[24] Sweeney, R.A., Generic combustion method for determination of crude protein in feeds: collaborative study. Journal-Association of Official Analytical Chemists, 1989. 72(5): p. 770.

[25] AOCS, Official methods and recommended practices of the American oil Chemists' Society, 1998, AOCS Press: Champaign.

[26] Roussel, S.A., et al., Detection of Roundup Ready (TM) soybeans by near-infrared spectroscopy. Appl Spectrosc, 2001. 55(10): p. 1425-1430.

[27] de Noord, O.N., Multivariate calibration standardization. Chemom. Intell. Lab. Sys, 1994. 25: p. 85-97.

Table 1. Summary statistics of the calibration and validation sets

Parameter	Instrument (serial)	N	Mean (%)	Range (%)	Standard deviation (% pts)
Protein (13% moisture basis)	Infratec 1229 (553075) Infrated 1241 (0350)	1101	36.27	24.72-46.50	3.42
	OmegAnalyzerG 106110 OmegAnalyzerG 106118	1100	36.25	24.72-46.89	3.41
	Validation set	154	34.92	28.93-43.87	3.48
	Standardization set	20	36.38	24.21-45.38	4.65
Oil (13% moisture basis)	Infratec 1229 (553075) Infrated 1241 (0350)	1123	18.37	12.48-24.64	1.91
	OmegAnalyzerG 106110 OmegAnalyzerG 106118	1118	18.30	11.85-24.64	1.92
	Validation set	155	18.51	13.43-22.95	2.00
	Standardization set	20	18.57	15.33-21.70	1.76

Table 2. Validation results of soybean protein calibration models of Bruins OmegAnalyzerGs (a) and Infratec Grain Analyzers (b)

(a)

Method	N_c[a]	N_v[b]	SEP[c] (% pts)	RPD[d]	Bias[e] (% pts)	LV[f]	Calibration	Validation
Original (106110)	122	154	0.53	6.55	-0.12	5	106110	106110
	403	154	0.47	7.38	0.07	5	106110	106110
	1100	154	0.47	7.39	0.18	9	106110	106110
Original (106118)	124	154	0.51	6.86	-0.63	3	106118	106118
	403	154	0.50	6.93	-0.07	4	106118	106118
	1100	154	0.49	7.11	0.07	4	106118	106118
No Standardization	122	154	0.58	6.00	-0.43	5	106110	106118
	403	154	0.52	6.75	-0.28	5	106110	106118
	1100	154	0.47	7.45	-0.32	5	106110	106118
Robust model	246	154	0.51	6.79	-0.62	4	6110+6118	106118
	806	154	0.50	6.95	-0.13	4	6110+6118	106118
	2200	154	0.47	7.34	-0.10	4	6110+6118	106118
Slope/Bias Correction	122	154	0.54	6.40	-0.43	5	106110	106118
	403	154	0.48	7.20	-0.11	5	106110	106118
	1100	154	0.50	7.02	-0.40	4	106110	106118
Standardization Set (106110)	122	20	0.57	8.17	-0.01	5	106110	STD[g]
	403	20	0.52	8.95	0.18	9	106110	STD
	1100	20	0.57	8.11	0.21	8	106110	STD
Standardization Set (106118)	124	19	0.55	8.43	-0.35	9	106118	STD
	403	19	0.61	7.81	-0.22	11	106118	STD
	1100	19	0.67	7.13	-0.05	12	106118	STD

[a] Number of samples in calibration set

[b] Number of samples in validation set

[c] Standard error of prediction

[d] SEP and Bias are expressed in % pts

[e] Relative predictive determinant

[f] Latent variable

[g] Standardization set

(b)

Method	N_c[a]	N_v[b]	SEP[c] (% pts)	RPD[d]	Bias[e] (% pts)	LV	Calibration	Validation
Original (1241(0350))	123	154	0.54	6.39	0.00	9	1241	1241
	405	154	0.51	6.83	0.13	9	1241	1241
	1101	154	0.47	7.46	0.19	10	1241	1241
Original (1229 (553075))	126	154	0.71	4.89	0.20	5	1229	1229
	405	154	0.62	5.59	0.22	10	1229	1229
	1101	154	0.55	6.29	0.32	7	1229	1229
No Standardization	123	154	0.61	5.73	0.01	9	1241	1229
	405	154	0.61	5.71	0.21	12	1241	1229
	1101	154	0.58	6.05	0.34	9	1241	1229
Robust model	249	154	0.62	5.58	0.06	3	1241+1229	1229
	810	154	0.57	6.16	0.27	3	1241+1229	1229
	2202	154	0.55	6.33	0.39	4	1241+1229	1229
Slope/Bias Correction	123	154	0.59	5.85	0.11	6	1241	1229
	405	154	0.55	6.32	0.33	3	1241	1229
	1101	154	0.54	6.42	0.35	4	1241	1229
Standardization Set (1241(0350))	123	20	0.20	21.85	-0.01	4	1241	STD
	405	20	0.20	22.51	-0.09	7	1241	STD
	1101	20	0.21	20.60	0.01	7	1241	STD
Standardization Set (1229 (553075))	126	20	0.23	19.37	-0.07	7	1229	STD
	405	20	0.18	24.22	-0.10	8	1229	STD
	1101	20	0.17	25.92	0.05	7	1229	STD

Table 3. Validation results of soybean oil calibration models of Bruins OmegAnalyzerGs (a) and Infratec Grain Analyzer (b)

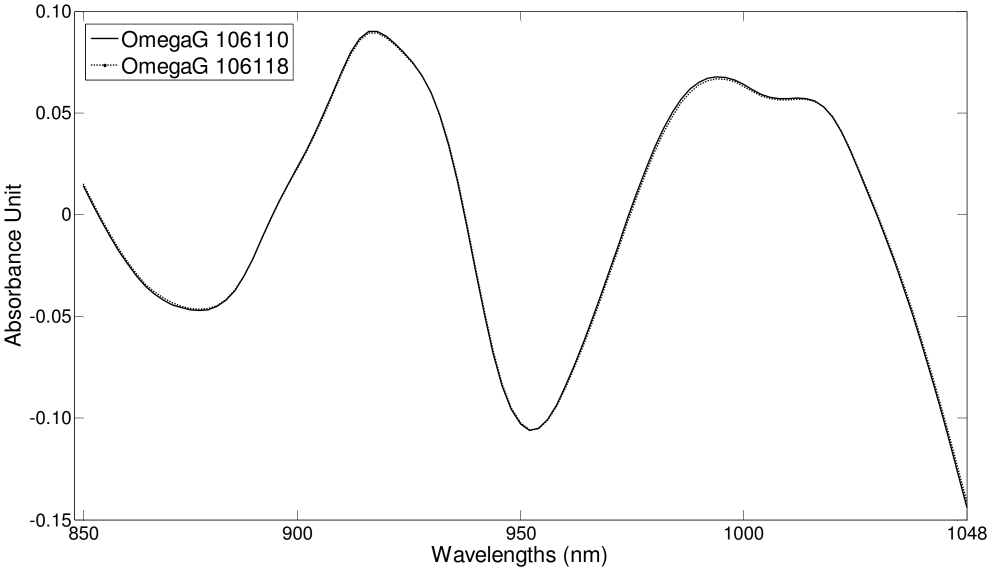
(a)

Method	N_c[a]	N_v[b]	SEP[c] (% pts)	RPD[d]	Bias[e] (% pts)	LV	Calibration	Validation
Original (106110)	117	155	0.39	5.17	-0.29	6	106110	106110
	372	155	0.37	5.47	-0.32	9	106110	106110
	1118	155	0.37	5.44	-0.39	5	106110	106110
Original (106118)	117	155	0.36	5.52	0.01	6	106118	106118
	372	155	0.36	5.63	-0.13	7	106118	106118
	1118	155	0.34	5.97	-0.12	7	106118	106118
No Standardization	117	155	0.36	5.52	0.01	6	106110	106118
	372	155	0.36	5.63	-0.13	7	106110	106118
	1118	155	0.34	5.97	-0.12	7	106110	106118
Robust model	234	155	0.35	5.76	-0.13	6	6110+6118	106118
	744	155	0.35	5.77	-0.09	6	6110+6118	106118
	2236	155	0.34	5.91	-0.10	6	6110+6118	106118
Slope/Bias Correction	117	155	0.34	5.88	-0.49	6	106110	106118
	372	155	0.34	5.83	-0.48	7	106110	106118
	1118	155	0.34	5.94	-0.43	7	106110	106118
Standardization Set (106110)	117	20	0.22	8.12	-0.27	5	106110	STD
	372	20	0.21	8.42	-0.22	8	106110	STD
	1118	20	0.21	8.23	-0.22	7	106110	STD
Standardization Set (106118)	117	19	0.18	9.97	0.07	7	106118	STD
	372	19	0.18	9.83	0.01	6	106118	STD
	1118	19	0.18	9.96	-0.01	5	106118	STD

(b)

Method	N_c[a]	N_v[b]	SEP[c] (% pts)	RPD[d]	Bias[e] (% pts)	LV	Calibration	Validation
Original (1241(0350))	116	155	0.35	5.68	-0.30	8	1241	1241
	373	155	0.34	5.81	-0.40	9	1241	1241
	1123	155	0.33	6.07	-0.48	12	1241	1241
Original (1229 (553075))	116	155	0.36	5.49	-0.25	8	1229	1229
	373	155	0.34	5.81	-0.18	7	1229	1229
	1123	155	0.33	6.10	-0.25	7	1229	1229
No Standardization	116	155	0.38	5.21	-0.41	6	1241	1229
	373	155	0.36	5.56	-0.48	6	1241	1229
	1123	155	`0.35	5.65	-0.56	5	1241	1229
Robust model	232	155	0.36	5.60	-0.30	6	1241+1229	1229
	746	155	0.35	5.79	-0.39	6	1241+1229	1229
	2236	155	0.34	5.97	-0.47	8	1241+1229	1229
Slope/Bias Correction	116	155	0.36	5.55	-0.76	6	1241	1229
	373	155	0.35	5.65	-0.72	5	1241	1229
	1123	155	0.34	5.91	-0.89	5	1241	1229
Standardization Set (1241(0350))	116	20	0.23	5.72	-0.51	9	1241	STD
	373	20	0.21	7.75	-0.55	10	1241	STD
	1123	20	0.20	8.43	-0.45	10	1241	STD
Standardization Set (1229 (553075))	116	20	0.18	9.35	-0.18	11	1229	STD
	373	20	0.19	8.89	-0.13	10	1229	STD
	1123	20	0.19	8.89	-0.13	10	1229	STD

(a)



(b)

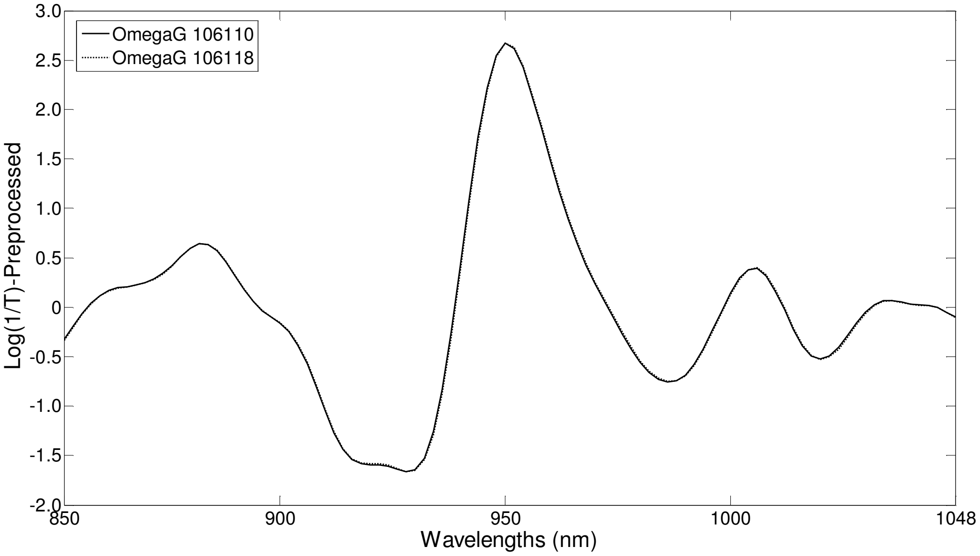
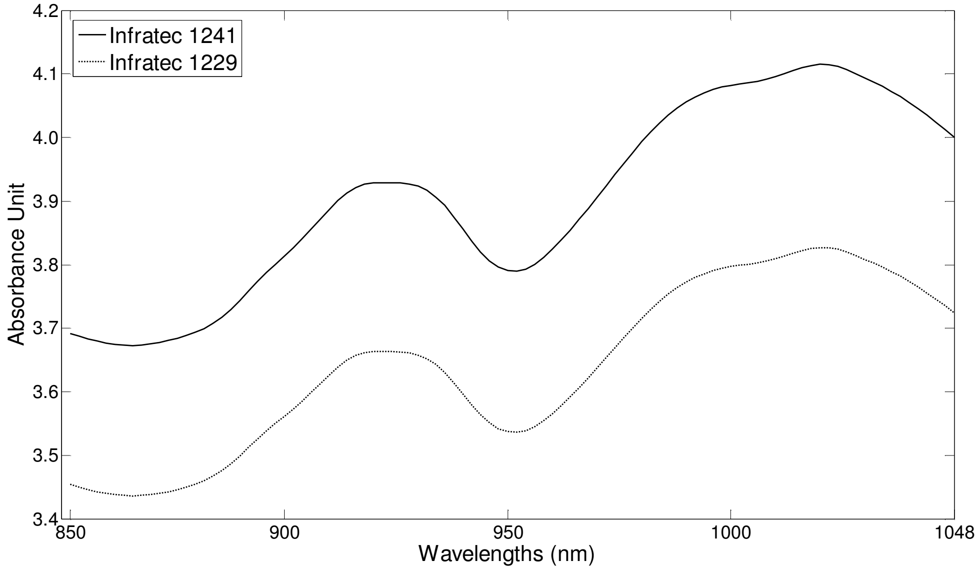


Figure 1. Averaged raw spectra of all calibration samples (a), preprocessed spectra with second derivative, SNV and detrending (b) of Bruins OmegAnalyzerGs

(a)



(b)

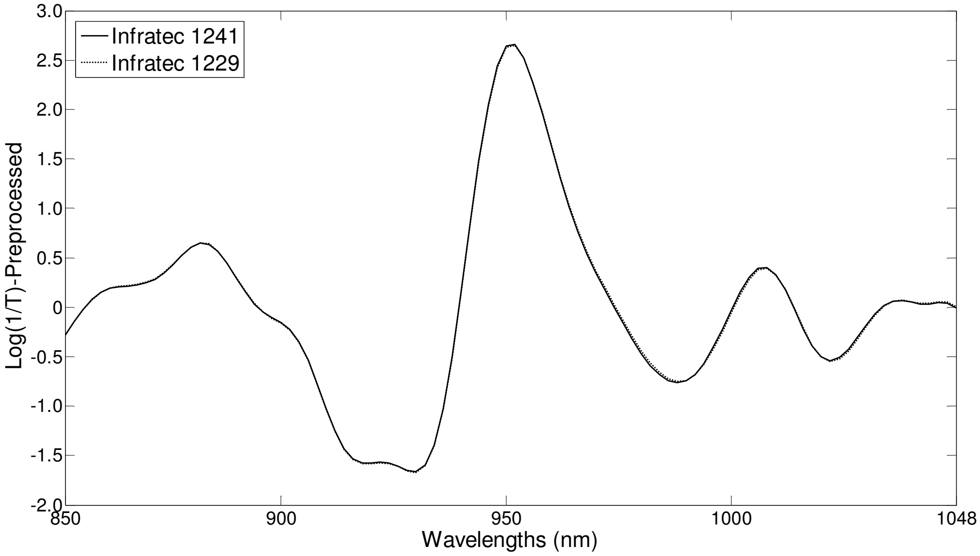
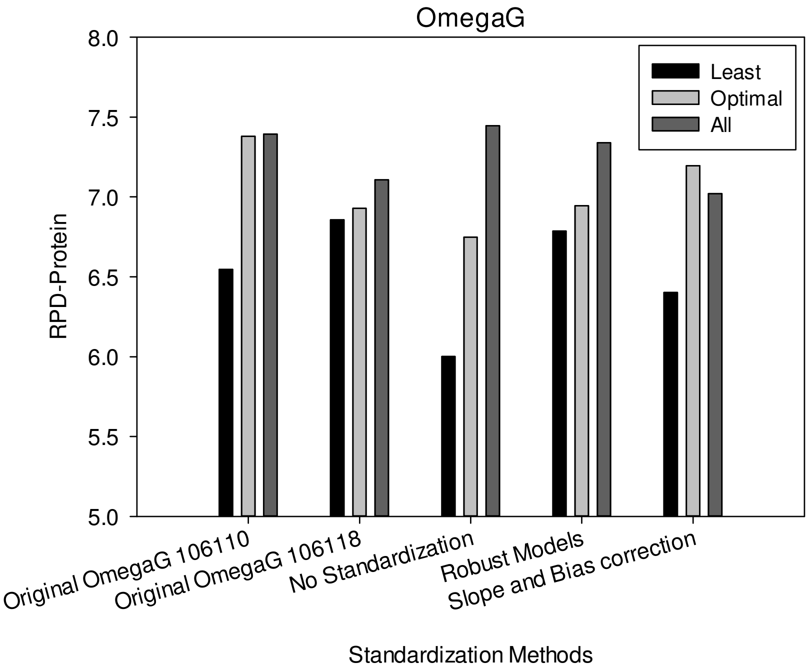


Figure 2. Averaged raw spectra of all calibration samples (a), preprocessed spectra with second derivative, SNV and detrending (b) of Infratec Grain Analyzers

(a)



(b)

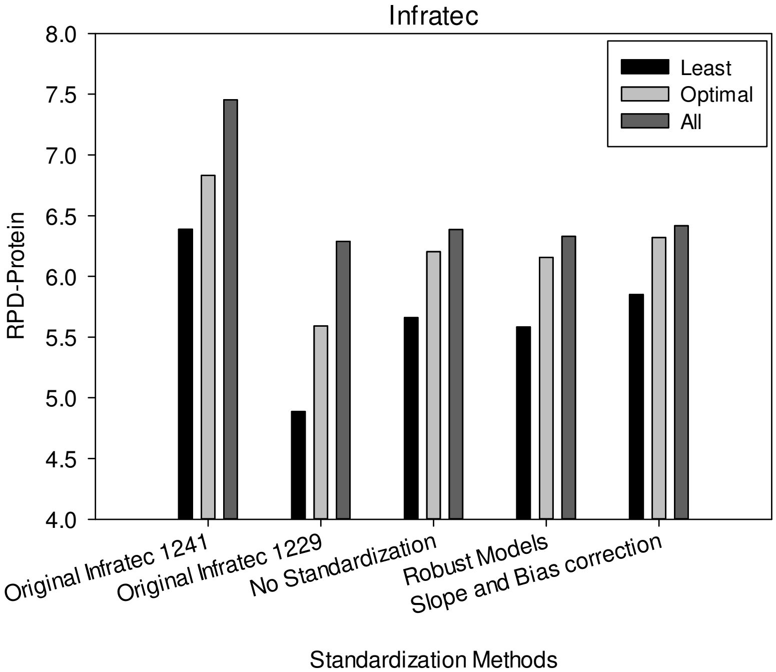
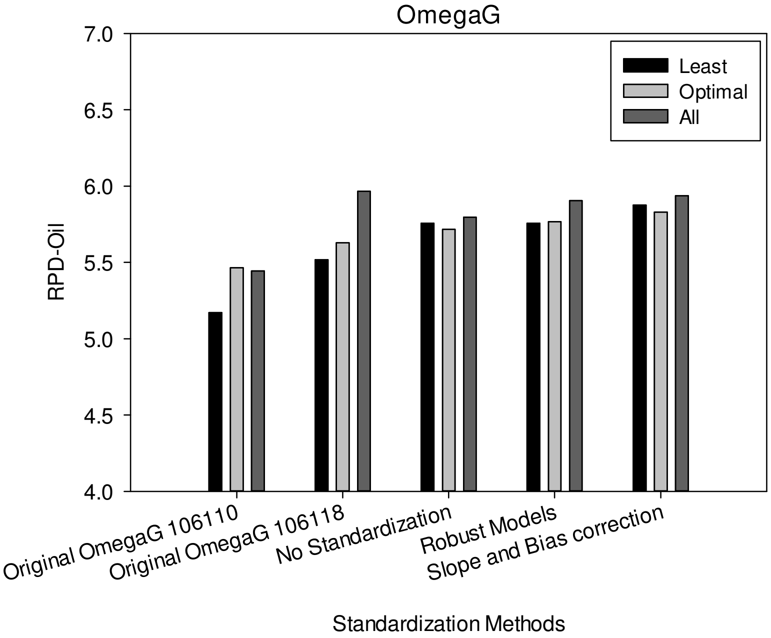


Figure 3. Validation results for soybean protein with different calibration subsets and standardization methods on Bruins OmegAnalyzerGs (a) and Infratec Grain Analyzers (b)

(a)



(b)

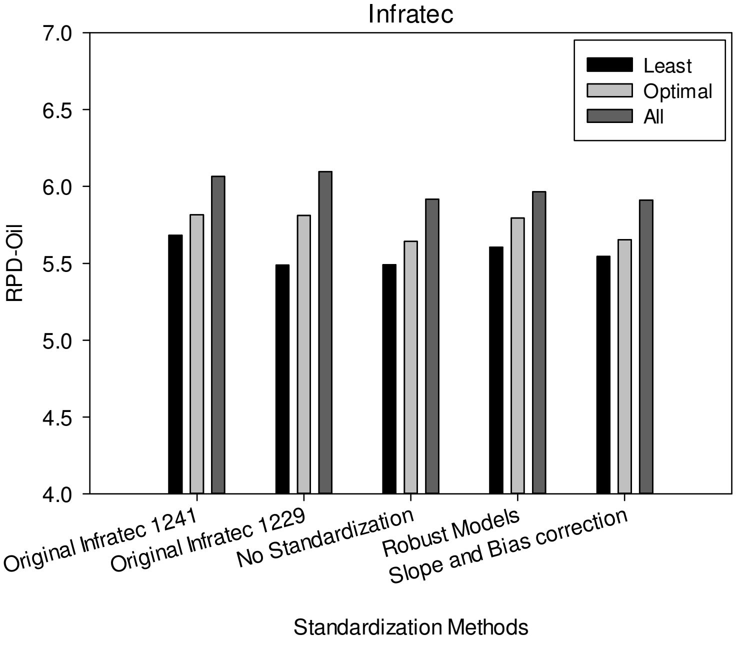


Figure 4. Validation results for soybean oil with different calibration subsets and standardization methods on Bruins OmegAnalyzerGs (a) and Infratec Grain Analyzers (b)

CHAPTER 4. ROBUST TEMPERATURE COMPENSATION FOR NEAR-INFRARED TRANSMITTANCE CALIBRATION OF PROTEIN AND OIL IN SOYBEANS

A paper to be submitted to the Journal of Agricultural and Food Chemistry

Nanning Cao[a], Charles R. Hurburgh[a]

[a] Iowa State University, Agricultural and Biosystems Engineering, Ames, Iowa 50010, USA.

Abstract

Near infrared (NIR) technology has been applied widely to provide fast analysis on grain quality. In practical application, temperature fluctuation occurs during grain handling process after harvest, due to the changes of weather and storage conditions. This paper assessed the effect of soybean sample temperature on the prediction performance of NIR calibrations and the temperature compensation methods. Three types of models (extended global model, simulated global model and difference augmentation) with samples scanned at different temperatures (5, 22 and 45 °C) were constructed. Extended global model directly included five temperature samples into calibration set. Simulated global model was built on spectra simulated from room temperature samples and 10 difference spectra. Difference augmentation method added the generated difference temperature spectra as a noise simulation to the original spectra. These models were compared with a local model built on samples collected at room temperature (22 °C). These compensated models aim at decreasing the prediction errors of protein and oil contents in soybeans. A small number of samples ($n \approx 120$) in addition to five representative temperature samples were used in calibration set. The extended global model and difference augmentation method gave similar results on the predictions of temperature set with decreased SEPs of 0.60% and 0.47% for protein and oil, respectively, and 0.58% and 0.46%, respectively on Infratec. With temperature compensated models, the prediction errors on regular samples measured at room temperature were also reduced from 0.53% to 0.51% for protein and from 0.39% to 0.34% for oil on Bruins, and from 0.54% to 0.52% for protein and from 0.35% to 0.34% for oil on Infratec as small fluctuations in temperature were corrected.

KEYWORDS: Temperature compensation; Robust calibration; Near-infrared spectroscopy; Soybean

Introduction

Near-Infrared spectroscopy (NIRS) has been successfully utilized in grain quality measurement for quality assurance[1-3]. It provides real time rapid testing of samples. Soybean [*Glycine max*] is a major source of plant protein and oil. Several characteristics of soybean and its derived foods are attributable to soybean protein and oil contents. In NIRS analysis, robust calibration is critical for accurate chemical determinations. In order to obtain an effective calibration, the calibration data need to be comprehensive and cover all types of variations that attract the factor of interest. In the real-world application of calibration, there are still problems that make it difficult in long-term practical use. One such problem is the environmental and instrumental temperature fluctuation during measurements. The ambient temperature is an environmental factor that has a considerable influence on NIR spectra.

Agricultural and food products are complex materials made up of carbohydrate, protein, fibrous components and water. Water has an important impact on the NIR spectra due to its strong absorption. The NIR spectrum of water is very sensitive to temperature. This is because the weak forces that influence the molecular bonding are easily affected by the change of temperature.

There are several approaches to deal with temperature effect in agricultural products. In the early 1980s, temperature was proven responsible for the differences discovered between Federal Grain Inspection Service (USDA-FGIS) laboratories. Williams et al. examined the influence of temperature on protein and moisture in wheat[4]. A method was developed to correct the temperature of ground wheat samples to room temperature. The calibration model included samples with a range of temperatures. The temperature effect was not associated with significant slope change of the calibration equations. Thus, an intercept adjustment could be used. An inverse relationship was found between temperature of ground samples and their corresponding apparent protein contents. Currently, global model built with spectra collected at different temperatures was the most commonly used method[5-7]. The advantage of this method is easy operation, with no temperature information needed for prediction. Another approach is temperature compensation firstly proposed by Kawano et al.[8]. A temperature compensation factor was suggested by using a correlation chart to correct prediction of Brix values in peaches. Similar techniques were applied to build a universal calibration that was robust against temperature effect[9]. Selected wavelengths were used to estimate the effect of the sample temperature on bias. Bias could be reduced by selecting adequate wavelengths for calibration, which could not always achieved by MLR. This study gave mathematical and theoretical analysis of K-method and proved its applicability. The third method is to correct temperature shifts by calibration transfer method[10, 11]. With modified piecewise direct standardization (PDS), temperature variation was regarded as a special type of instrumental variation at different temperatures. Another example is a generalized PDS method proposed as continuous piecewise direct standardization (CPDS) were developed to deal with temperature as a discrete variable[12]. The fourth method is to remove temperature effect from spectral data by a preprocessing method. The external parameter orthogonalisation (EPO) was proposed to remove the information in spectra that are mostly influenced by external parameter variation[13], in this case temperature.

However, there are few studies on the compensation of temperature effect on calibration for the composition of soybean. The current whole grain transmission calibration procedure involves developing a base calibration followed by the addition of samples to the base calibration for instrument stabilization and temperature stabilization. Instrument temperature is usually controlled by heating and cooling circuitry within an instrument. Even with a temperature control module, small temperature fluctuation exists in practical condition. A robust calibration is able to provide precise predictions and not vulnerable to different perturbation factors and measurement noise in the control system[14]. In this case, it is necessary to stabilize sample temperature effects in NIRS calibrations and/or to test the stability of calibrations to sample temperature.

The objective was to create robust models against temperature perturbation with the least experimental effort and the lowest cost associated with data acquisition. This study assessed the influence of temperature on NIR transmittance spectra of soybeans and compared three strategies (extended global model, simulated global model and difference augmentation) for temperature compensation. For comparison, a local model built on spectra collected at room temperature only was used a benchmark.

Theory

Temperature effect on spectral data

In NIR spectra, a sample temperature rise will increase the probability of molecules transfer to a high energy level, which means more radiation is absorbed and less is reflected. The increased sample temperature raises the vibration energy between the molecules so that molecular bonds, especially the hydrogen bonds break. Accordingly, the clusters of the water molecules become smaller and the absorbance of the free hydrogen bonds increases. Likewise, the reflectance decreases with increasing temperature. The change in the temperature could be connected with the changes in the hydrogen bonding of water[15]. The band near 1449 nm is the temperature-dependent region of water.

The influence of sample temperature on its NIR spectra has been investigated in several kinds of food and agricultural products[6, 16]. With the fundamental knowledge of water spectra, it is known that large spectral variations of water caused by temperature change can attribute to changes of the hydrogen-bonded water structure. However, in complex mixtures like food and agricultural products, the temperature perturbation is likely to be more complex especially in high moisture content products. Thus, it is hard to summarize in one variable. Sometimes, temperature changes may even lead to phase change, which is more complex than changes within the same phase. In grain products, NIR spectra are highly responding to moisture content[17]. In this study, moisture content of soybean samples differs from 4% to 18%. In order to eliminate the effect of different moisture contents, reference values of both protein and oil were converted into a constant moisture basis (13%).

Temperature effect on calibration

External perturbation like temperature has a nonlinear impact on the spectral shape. The influence of temperature on calibration has been examined [18] by comparing the local model with no temperature variation and global model included samples at different temperatures. Global models were proved to perform equally well as local models calibrated at a specific temperature. Experimental design is usually applied to span the concentration variations with the least number of representative samples. However, this is not always possible on agricultural products. In some cases, it was achieved by adding artificial ingredients into the mixtures to create samples that match with experimental design points. Milk samples were simulated by adding whey protein and oil mixtures to form different concentrations of protein and fat [19]. This could not be applied to grain products like whole soybean sample. To solve this problem, the original calibration set contains representative samples of all concentration levels that were deliberately selected from a large population for both protein and oil contents.

Materials and Methods

Samples

Calibration dataset

The original calibration set included approximately 120 whole soybean samples with reference values of protein and oil collected and was scanned at Grain Quality Laboratory, Ames, IA, USA, during crop years 2001-2010. This calibration subset was selected from an original pool with more than 1,000 soybean samples. Samples in this set were run at room temperature (22 ± 2 °C).

Temperature dataset

Twenty samples selected from crop year 1994 to 2012 were included in the temperature set. Samples in this set were tested the National Type Evaluation Program (NTEP) process, which requires samples to be measured under room temperature (22 ± 2 °C), cold (5 ± 2 °C), room temperature (22 ± 2 °C), warm (45 ± 2 °C) and room temperature (22 ± 2 °C) conditions. For every temperature change, samples were equilibrated to the target temperature under room temperature, cooler or oven for 24 hours. Each sample was scanned five times. Sample temperature was measured in the hopper before scanning using the non-contact infrared thermometer.

Five representative samples spanning the reference variation were selected from the temperature set by Kennard-Stone algorithm[20] to generate temperature information for calibration set in global model and difference augmentation methods. The remaining fifteen samples were used for validation.

Validation datasets

Two datasets were used validation sets to test the performance of temperature correction. Validation set 1 was composed of 15 temperature samples scanned through NTEP process with 5 spectra for each sample at temperature 5, 22 and 45°C.

Soybean samples from crop year 2011 were used as an external validation set 2. This set was completely independent and was indicative of future prediction samples. No outlier detection process was applied to entire validation set. Samples in this set were all scanned at room temperature at 22 °C. This set was applied to test the robustness of calibration models after temperature correction under normal condition.

Calibration performance was evaluated in terms of precision, accuracy, and model fit. The standard error of prediction (SEP) on the validation sets or standard deviation of differences were used to evaluate the precision. SEP estimates the typical difference between prediction and reference values. The accuracy will be determined by the bias (average of differences). Bias is a good indicator of similarity between validation samples and the calibration set [21]. Root mean square error of prediction (RMSEP) is the total error and is equal to the quadrature addition of SEP and bias.

Reference chemical analyses

Protein content was determined by Eurofins (Eurofins Scientific, Inc., Des Moines, Iowa, USA) using the combustion method (AOAC 990.03)[22] and oil content using ether extract (AOCS Ac3-44)[23]. Eurofins did both analyses. Concentrations for protein and oil were converted to a 13% moisture basis. Summary statistics for the calibration, temperature set and validation sets are presented in Table 1.

Instrumentation

Soybean samples were scanned by a Bruins OmegAnalyzerG (serial number: 106110) (Bruins Instruments, Puchheim, Germany) and a Foss Infratec Grain Analyzer 1241 (serial number: 12410350) (FOSS North America, Eden Prairie, MN, USA). Both are transmittance instruments with effective spectral range from 850 to 1048 nm with a 2nm increment. A path length of 30 mm was used.

Data Analysis

The spectral data are analyzed on the region 850-1048 nm. Second derivative was applied to the raw spectra first. Then, we considered standard normal variate (SNV) followed by detrending (DT). As a preprocessing technique, SNV corrects for the linear baseline shift and signal intensity variations. It normalizes each spectrum to reduce scattering effects due to packing heterogeneity or path-length variations[24]. Detrending can be used after SNV to account for the variation in baseline shift and linearity of spectra by using a second-degree polynomial to correct each spectrum[25].

PLS regression was used to develop all prediction models. In PLS, the original matrix is compressed into latent variables (LVs) that maximize the covariance between the reference values and all possible linear functions of the spectral data. PLS has been a classical analytical technique to handle the multivariate natures of the agriculture products and the highly collinear NIR spectroscopic data. The number of PLS components for each model was selected by cross-validation.

Calculations were performed with MATLAB R2011a (The MathWorks, Natick, MA) installed with the PLS_Toolbox v.6.2.1 (Eigenvector Research, Wenatchee, WA). Kennard-Stone method was computed in a custom-created Matlab program.

Local models

The benchmark calibration models were constructed on spectra collected at room temperature (22 °C), then were used to predict samples (validation set 1) collected at different temperatures (5, 22 and 45 °C). The robustness of these calibration models were test in previous study. These were the benchmark model that was used to compare with temperature compensation methods. The prediction results of validation set 2 (samples scanned at room temperature) were also computed to test the predictive ability of the calibration models (benchmark and temperature controlled) on normal or unperturbed samples.

Global models

In this approach, calibration model was constructed on a dataset that contained spectra collected at different temperatures. This is a commonly used solution to model the effects of temperature. It only requires the inclusion of temperature samples in calibration set. The model is trained empirically to handle temperature as an unknown interference[6]. With one global model for samples analyzed at varying temperatures, it is not necessary to know the temperature of new prediction samples, nor the calibration set. A new calibration model is not needed for each temperature. The problem of this method is the large number of temperature samples in the calibration process, which increases the expense and effort of experiments, and its empirical (sample set dependent) nature.

In this study, two types of global models were created. The first one included only five well-chosen samples taken at all three temperatures in addition to the original calibration set. This is called the *extended global model*. Compared to the traditional global model with the inclusion of all the samples collected at all the temperatures, this kept the calibration model less complex and computational efficient.

The second one is called the *simulated global model*. In this approach, spectra collected at room temperature were simulated to spectra of cold and warm samples. The average values of difference spectra were used for simulation. The difference spectra were calculated as follow:

$$\mathbf{d} = \begin{cases} \mathbf{x}_w - \mathbf{x}_0 \; \textit{and} \\ \mathbf{x}_0 - \mathbf{x}_c \end{cases} \tag{62}$$

Spectra of cold ($\mathbf{x}_c$) and warm ($\mathbf{x}_w$) samples were simulated by

$$\begin{cases} \mathbf{x}_c = \mathbf{x}_0 - \overline{\mathbf{d}_c} \\ \mathbf{x}_w = \mathbf{x}_0 + \overline{\mathbf{d}_w} \end{cases} \tag{63}$$

where $\mathbf{x}_w, \mathbf{x}_c$ represent spectra analyzed at warm and cold conditions, respectively.

Then, the simulated spectral matrix $\mathbf{X}_{\text{sim}}$ and response vector ($\mathbf{y}_{\text{sim}}$) would be:

$$\mathbf{X}_{\text{sim}} = \begin{pmatrix} \mathbf{X}_c \\ \mathbf{X}_R \\ \mathbf{X}_W \end{pmatrix} \text{ and } \mathbf{y}_{\text{sim}} = \begin{pmatrix} \mathbf{y} \\ \vdots \\ \mathbf{y} \end{pmatrix}$$

Difference augmentation method

Segtnan et al. proposed the difference augmentation method to simulate the effect of temperature on spectra [5, 26]. This is an ensemble method that based on the difference spectra for temperature correction. The purpose is to augment the temperature effect in the spectral variation in calibration set in order to handle the drift expected in future samples.

A subset of 5 representative samples was selected from the temperature set that were run at three levels of temperatures (5, 22, 45 °C). For grain quality measurement, the middle temperature (25 °C) is the expected normal condition in future analysis. In this case, spectra scanned at room temperature (25 °C) was used as basis and defined as $\mathbf{x_0}$. The difference spectra were calculated in equation (1).

In this study, five well-chosen representative samples collected at three different temperatures (5, 22, 45 °C) were used as temperature samples to generate 10 difference spectra. With each difference spectra, original calibration samples were simulated to perturbed spectra:

$$\mathbf{x_{aug}} = \mathbf{x_0} + r\mathbf{d} \tag{64}$$

where r represents n (number of samples in calibration set) independent Gaussian random numbers with standard deviation 0.5. Then, 10 matrices were generated to form the simulated augmentation calibration set. The final matrix used for calibration was composed by stacking these 10 matrices. The corresponding y values were stacked by 10 copies of the original reference values. Details of this method are described in reference[5].

Results and Discussion

Local models

Local models were used as benchmark calibration models for comparison with the temperature compensation methods. Local models were used to predict protein content in soybean samples analyzed at 5 °C (cold) and 45 °C (warm). The regression results were plotted in Figure 3 for Bruins 106110 (a) and Infratec 1241 (b). Local calibration models tended to overestimate warm samples and underestimate cold samples on Bruins 106110. For both protein and oil contents, local models gave the largest SEPs and bias. Large bias indicated uncontrolled variables in prediction samples. In this case, uncontrolled temperature variable affects the predictive ability of calibration models when samples taken at temperatures present. It was considered water was an important factor causing a bias due to a variation of temperature because the water absorption was easily affected by it [8]. The reason of converting to 13% moisture basis is that temperature compensation method worked better with constant moisture content. By contrast, Infratec 1241 models were less sensitive to temperature perturbation in sample spectra in terms of both protein and oil contents. They gave reasonable SEPs for both protein and oil. However, the bias of protein local model was relatively large.

Extended global models

In addition to the original calibration set, five temperature samples were also included in the extended global model. For Bruins 106110 models, both SEPs and bias for protein and oil were decreased on validation set 1 and 2. This implies that with only five representative samples, extended global models included the temperature information to give better prediction and improved the robustness of calibration models. For Infratec 1241, extended global models improved the prediction performance on validation set 1 and gave similar or slightly better results on validation set 2.

Simulated global models

This approach is a simulation of the full global model. To visualize the temperature effects in the soybean spectra, plots in Figure 1 gave an example on the similarity of the simulated temperature spectra and the ones taken at the real temperatures (a). The simulated spectra were plotted in Figure 1(b). By adding the difference spectra to spectra scanned at room temperature, the simulated cold and warm spectra were assemble the real spectra. This indicated the use of simulation based on five temperature samples artificially created spectra resembling ones taken at varying temperatures. Compared to the local models, this method successfully decreased SEPs in prediction on both validation sets by involving temperature information. However, the problem of this approach is that the bias tends to be relatively large compared to other compensation methods. The reason is probably that in order to simulate a full global model, the average difference spectrum needs to be representative, which is very hard for complex mixtures as agricultural products. The effect of temperature on individual components in soybean may interfere with each other, which makes the simulation difficult.

Difference augmentation

The difference augmentation method worked well and obtained the lowest SEPs in most cases. By adding each difference spectrum as simulated temperature information on each spectrum in calibration set, the final augmentation matrix modeled temperature as an external perturbation factor. Consequently, difference augmentation method provided calibrations computed on a small calibration set and only five representative temperature samples to save the cost on acquirement of spectra and reference values.

Discussion

Room temperature (22 ± 2 °C) was used as a basis for comparison between the different methods investigated. The performance of these four methods is summarized Table 2 for both protein and oil contents. Validation set 1includes spectra of 15 spectra, each of which was scanned 5 times at three different temperatures. The prediction results show that the three compensation methods improved the prediction performance while taking the temperature effects into account. Moreover, these temperature models maintained or even improved their predictive ability on validation set 2, which was not perturbed by temperature changes. This may be because that the room temperature is not always constant during each run. On the other hand, the instrument temperature may fluctuate even with a temperature control function. The greater temperature variations of the compensated models are likely to give more stable temperature correction than the small and uncontrolled temperature variation of spectra taken at the room temperature. Temperature compensation improves the robustness of the calibration models, even under "normal" conditions, in which temperature is not controlled.

Figure 1(a) shows real spectra collected at different temperature. It is obvious that the absorbance at water peak around 980 nm increased with higher temperature. Wavelength 980 nm corresponds to the second overtone of O-H bond. Increased vibration energy among the molecules over elevated temperatures created higher absorbance of the free hydrogen bonds.

Peak shifts were observed in Figure 1(a). For this specific sample, the water peak 986 nm at room temperature shifted to 992 nm at 5 °C and 984nm at 45 °C. As temperature rises, the water peak shifts a few nanometer to shorter wavelength (lower frequency) with strengthening hydrogen bonding due to shifts from low density water (increasing expanded structure) to high density water (increasing collapsed structure). Temperature changes may also induce the overlapping of spectra of individual components in the complex mixture.

Principal component analysis (PCA) was conducted on spectra measured in Bruins 106110 and the loading plot of factor is shown in Figure 2 (a). Principal components (PCs) express the main variations of the spectra, while PC 1 explained 78.17% of the total variation in this case. The positive peak frequency in the loading plot at 968 nm corresponds to the absorption band of hydrogen bond. Difference spectra of soybean sample calculated by subtracting the spectrum at 22 °C were plotted in Figure 2 (b). Peaks appeared at 974 nm and 976 nm and become stronger with higher temperature, very close to the peak position in the loading plot. When the noise and baseline change are small on Bruins 106110, the temperature effect on water spectra is very regular[15]. This is reasonable since the broad water peak dominates in the region of 960-990 nm and is very sensitive to temperature changes. Loading plot of spectra collected on Infratec 1241 was not displayed due to the presence of obvious baseline shift.

The prediction residuals of protein content on temperature validation set were plotted in Figure 4. For Bruins 106110 (left column), the residuals of local model were scattered. Warm samples had positive residuals because local model overestimated protein contents for samples with higher temperatures. Residuals significantly decreased and no systematic bias remained after temperature compensation methods were applied to calibration models in all cases. For Infratec 1241 (right column), changes in residual distributions are not significant due to the robustness of its local model.

Calibration models with temperature compensation methods used one or two more latent variables than the local model. This is reasonable since the impact of temperature on spectra is not linear. Consequently, a linear regression method like PLS tends to use more regression factors. It also indicates these methods did not increase the complexity of calibration models significantly.

Conclusions

In this study, it has been shown that regarding the temperature effect on NIR transmittance spectroscopy, the robustness of calibration model could be improved by temperature compensation methods. Extended global model directly included five temperature samples into calibration set. Simulated global model was built on spectra simulated from room temperature samples and 10 difference spectra. Difference augmentation method added the generated difference temperature spectra as a noise simulation to the original spectra. All three temperature compensation methods provided improvements on the precision of prediction on both temperature set and regular validation set. Difference augmentation method reached the lowest SEPs in temperature set for both instruments. Bruins models were more sensitive to temperature effect. With difference augmentation method, SEPs of temperature sample set were reduced from 0.78% to 0.60% for protein and from 0.52% to 0.47% for oil on Bruins, and from 0.60% to 0.58% for protein and from 0.50% to 0.46% for oil on Infratec. Regarding regular validation samples scanned at room temperature, SEPs were also reduced from 0.53% to 0.51% for protein and from 0.39% to 0.34% for oil on Bruins, and from 0.54% to 0.52% for protein and from 0.35% to 0.34% for oil on Infratec.

In addition to the original calibration samples ($n \approx 120$), five well-selected samples run at three different temperatures provided sufficient information in calibration. Smaller number of calibration and temperature samples significantly reduced the cost of spectral measurements and the complexity of model. With a few more PCs, the complexity of model did not increase significantly due to the inclusion of temperature information. No temperature information is required for prediction samples, which provides huge advantage on future application.

References

1. Osborne, B.G., Recent developments in NIR analysis of grains and grain products. Cereal Foods World, 2000. 45(1): p. 11-15.

2. Kovalenko, I.V., G.R. Rippke, and C.R. Hurburgh, Determination of amino acid composition of soybeans (Glycine max) by near-infrared spectroscopy. Journal of agricultural and food chemistry, 2006. 54(10): p. 3485-3491.

3. Igne, B., G.R. Rippke, and C.R. Hurburgh, Measurement of Whole Soybean Fatty Acids by Near Infrared Spectroscopy. Journal of the American Oil Chemists Society, 2008. 85(12): p. 1105-1113.

4. Williams, P., K. Norris, and W. Zarowski, Influence of temperature on estimation of protein and moisture in wheat by near-infrared reflectance. Cereal Chemistry, 1982. 59(6): p. 473-477.

5. Segtnan, V.H., et al., Low-cost approaches to robust temperature compensation in near-infrared calibration and prediction situations. Applied spectroscopy, 2005. 59(6): p. 816-825.

6. Peirs, A., N. Scheerlinck, and B.M. Nicolaï, Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. Postharvest Biology and Technology, 2003. 30(3): p. 233-248.

7. Wulfert, F., et al., Linear techniques to correct for temperature-induced spectral variation in multivariate calibration. Chemometrics and intelligent laboratory systems, 2000. 51(2): p. 189-200.

8. Kawano, S., H. Abe, and M. Iwamoto, Development of a calibration equation with temperature compensation for determining the Brix value in intact peaches. Journal of Near Infrared Spectroscopy, 1995. 3: p. 211-218.

9. Abe, H., C. Iyo, and S. Kawano, A study on the universality of a calibration with sample temperature compensation. Journal of Near Infrared Spectroscopy, 2000. 8(3): p. 209-213.

10. Wang, Y.D. and B.R. Kowalski, Temperature compensating calibration transfer for nir filter instruments. Analytical Chemistry, 1993. 65: p. 1301-1303.

11. Barboza, F.D. and R.J. Poppi, Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature correction by transfer calibration procedures. Analytical and bioanalytical chemistry, 2003. 377(4): p. 695-701.

12. Wulfert, F., et al., Correction of temperature-induced spectral variation by continuous piecewise direct standardization. Analytical Chemistry, 2000. 72(7): p. 1639-1644.

13. Roger, J.M., F. Chauchard, and V. Bellon-Maurel, EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. Chemometrics and Intelligent Laboratory Systems, 2003. 66(2): p. 191-204.

14. Fernández Pierna, J.A., et al., How to build a robust model against perturbation factors with only a few reference values: A chemometric challenge at 'Chimiométrie 2007'. Chemometrics and Intelligent Laboratory Systems, 2011. 106(2): p. 152-159.

15. Maeda, H., et al., Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. Journal of Near Infrared Spectroscopy, 1995. 3: p. 191-202.

16. Büning-Pfaue, H., Analysis of water in food by near infrared spectroscopy. Food Chemistry, 2003. 82(1): p. 107-115.

17. Gergely, S. and A. Salgo, Changes in moisture content during wheat maturation-what is measured by near infrared spectroscopy? Journal of near infrared spectroscopy, 2003. 11(1): p. 17-26.

18. Wulfert, F., W.T. Kok, and A.K. Smilde, Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. Analytical Chemistry, 1998. 70(9): p. 1761-1767.

19. Zhang, X., et al., Influence and correction of temperature on optical measurement for fat and protein contents in a complex food model system. Infrared Physics & Technology, 2010. 53(3): p. 177-181.

20. Kennard, R.W. and L.A. Stone, Computer aided design of experiments. Technometrics, 1969: p. 137-148.

21. Kalivas, J.H., et al., Calibration Maintenance and Transfer Using Tikhonov Regularization Approaches. Appl Spectrosc, 2009. 63(7): p. 800-809.

22. Sweeney, R.A., Generic combustion method for determination of crude protein in feeds: collaborative study. Journal-Association of Official Analytical Chemists, 1989. 72(5): p. 770.

23. AOCS, Official methods and recommended practices of the American oil Chemists' Society, 1998, AOCS Press: Champaign.

24. Roussel, S.A., et al., Detection of Roundup Ready (TM) soybeans by near-infrared spectroscopy. Appl Spectrosc, 2001. 55(10): p. 1425-1430.

25. Verboven, S., M. Hubert, and P. Goos, Robust preprocessing and model selection for spectral data. Journal of Chemometrics, 2012. 26(6): p. 282-289.

26. Sáiz-Abajo, M.J., et al., Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. Analytica Chimica Acta, 2005. 533(2): p. 147-159.

Table 1. Summary statistics of the calibration, temperature and validation sets

Parameter	Instrument (serial)	N	Mean (%)	Range (%)	Standard deviation (% pts)
Protein (13% moisture basis)	Infrated 1241 (0350)	123	36.08	24.72-46.50	6.02
	OmegAnalyzerG 106110	122	36.25	24.72-46.89	6.17
	Temperature set	20	36.03	32.57-42.16	2.73
	Validation set 2	154	34.92	28.93-43.87	3.48
Oil (13% moisture basis)	Infrated 1241 (0350)	116	18.60	12.48-24.64	3.31
	OmegAnalyzerG 106110	117	18.56	11.85-24.64	3.38
	Temperature set	20	18.63	14.77-21.47	1.45
	Validation set 2	155	18.51	13.43-22.95	2.00

Table 2. Validation results of soybean calibration models of Bruins OmegAnalyzerG 106110 (a) and Infratec Grain Analyzer 1241 (b)

(a)

Constituent	Method	N_c[a]	Validation set 1 (Temperature set)				Validation set 2			
			SEP[b] (% pts)	Bias[c] (% pts)	RMSEP[d] (% pts)	LV[e]	SEP (% pts)	Bias (% pts)	RMSEP (% pts)	LV
Protein (13% moisture basis)	Local	122	0.78	-0.21	0.81	6	0.53	-0.12	0.54	5
	Extended global	147	0.62	-0.07	0.62	8	0.51	-0.02	0.51	5
	Simulated global	122*3	0.60	-0.40	0.72	7	0.53	-0.27	0.59	6
	Difference Augmentation	122+5	0.60	-0.28	0.66	7	0.51	-0.09	0.52	6
Oil (13% moisture basis)	Local	117	0.52	-0.13	0.53	5	0.39	-0.29	0.49	6
	Extended global	142	0.47	-0.05	0.47	8	0.35	-0.01	0.35	7
	Simulated global	117*3	0.47	-0.20	0.51	7	0.36	-0.12	0.38	7
	Difference Augmentation	117+5	0.47	-0.07	0.48	6	0.34	-0.03	0.34	7

[a] Number of samples in calibration set

[b] Standard error of prediction

[c] SEP and Bias are expressed in % pts

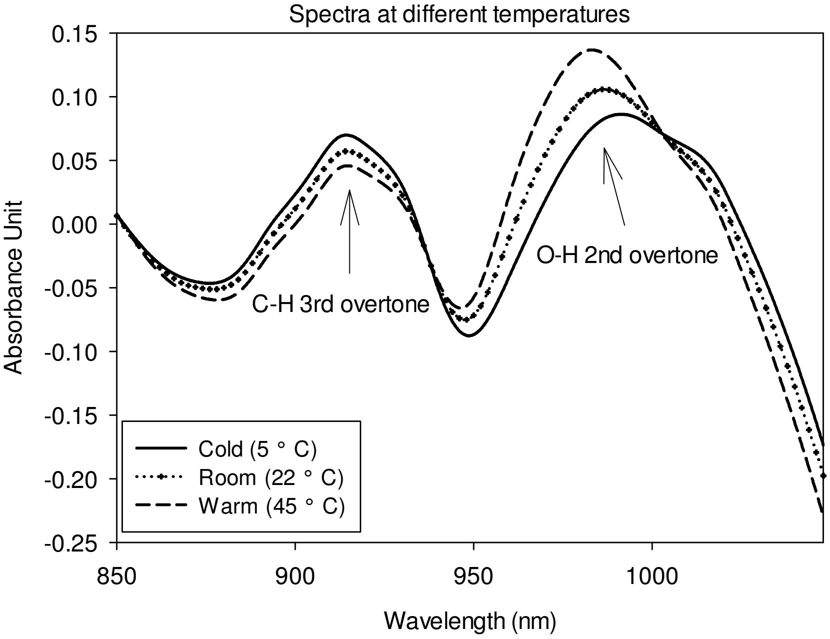
[d] Root mean square error of prediction

[e] Latent variable

(b)

Constituent	Method	N_c	Validation set 1 (Temperature set)				Validation set 2			
			SEP (% pts)	Bias (% pts)	RMSEP (% pts)	LV	SEP (% pts)	Bias (% pts)	RMSEP (% pts)	LV
Protein (13% moisture basis)	Local	123	0.60	-0.31	0.67	6	0.54	0.00	0.54	9
	Extended global	148	0.59	-0.19	0.62	8	0.55	-0.02	0.55	5
	Simulated global	123*3	0.57	-0.38	0.69	8	0.52	-0.08	0.52	5
	Difference Augmentation	123+5	0.58	-0.29	0.65	8	0.52	-0.12	0.53	5
Oil (13% moisture basis)	Local	116	0.50	0.02	0.50	5	0.35	-0.30	0.46	8
	Extended global	141	0.47	0.03	0.47	8	0.34	-0.09	0.36	6
	Simulated global	116*3	0.47	-0.11	0.49	7	0.34	-0.33	0.48	7
	Difference Augmentation	116+5	0.46	-0.08	0.47	8	0.35	-0.28	0.45	7

(a)



(b)

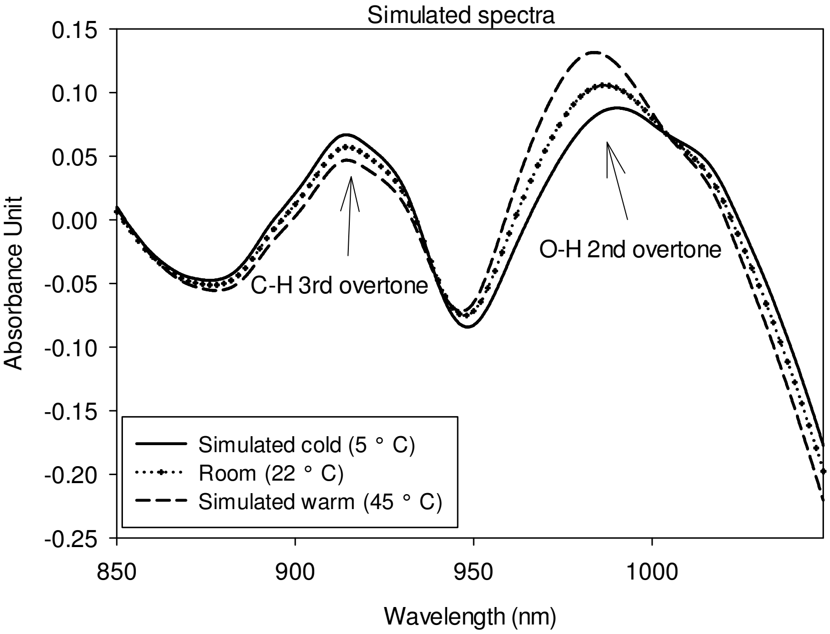
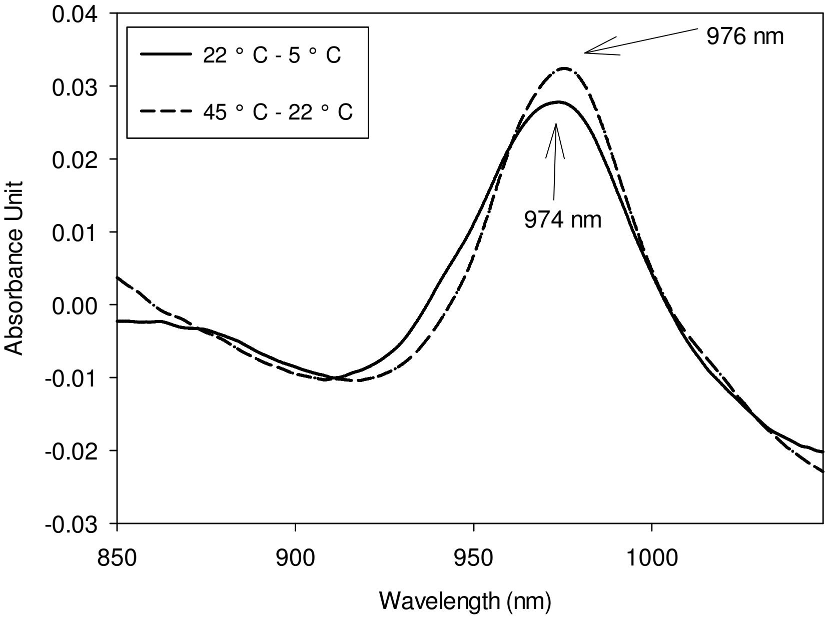


Figure 1. Spectra of soybean scanned at different temperatures (a) and simulated spectra (b) from spectra collected at room temperature on Bruins OmegAnalyzerG 106110

(a)



(b)

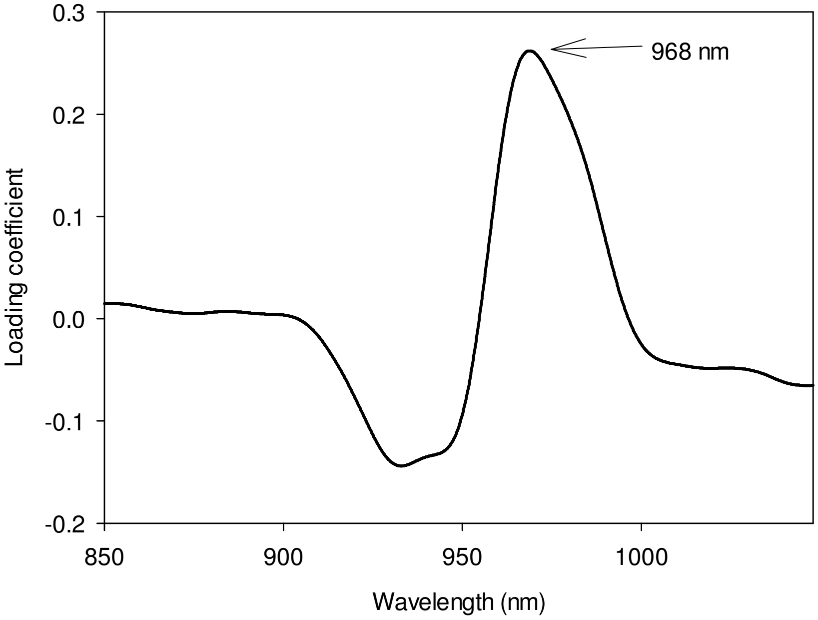
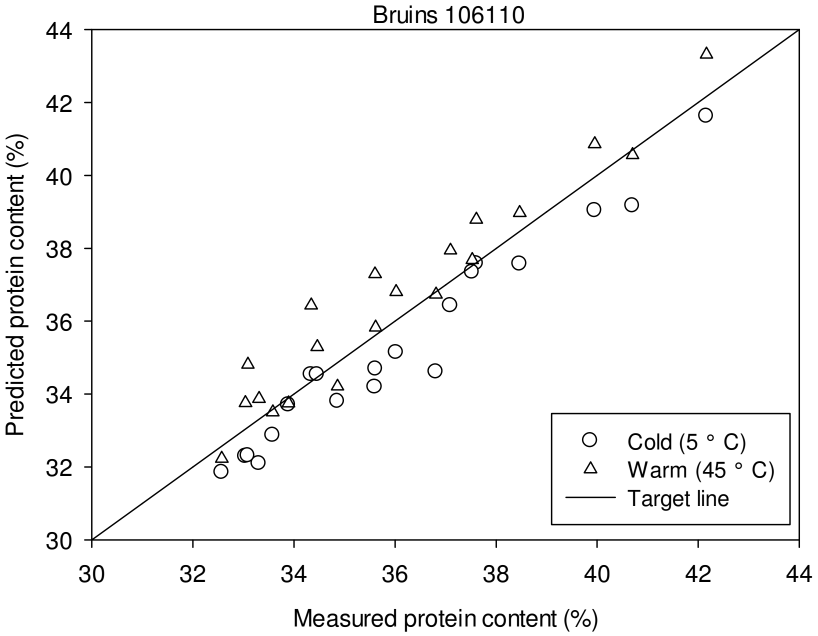


Figure 2. Difference spectra of soybean (a) by subtracting the spectrum at 22°C and PCA loading plot (b) of factor 1 for the model based on spectra measured at three temperatures on Bruins 106110

(a)



(b)

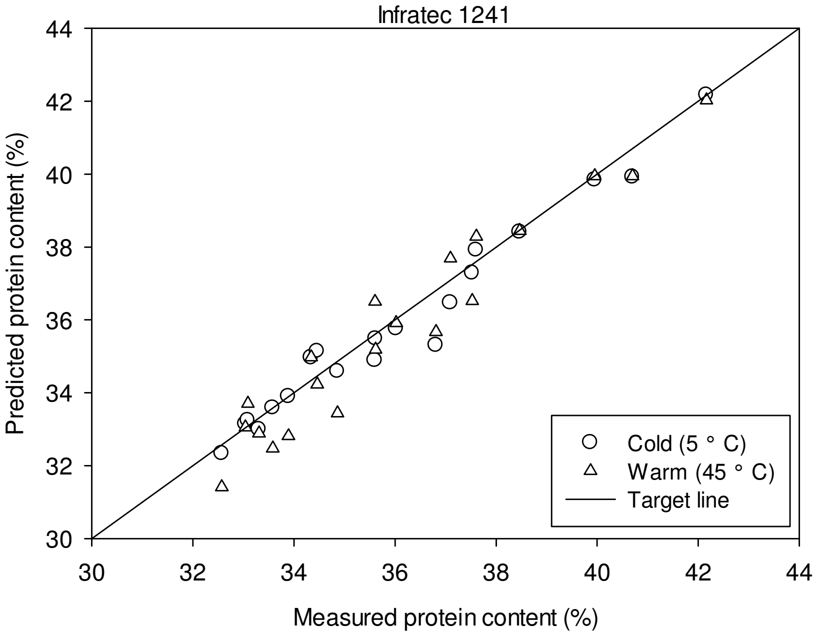
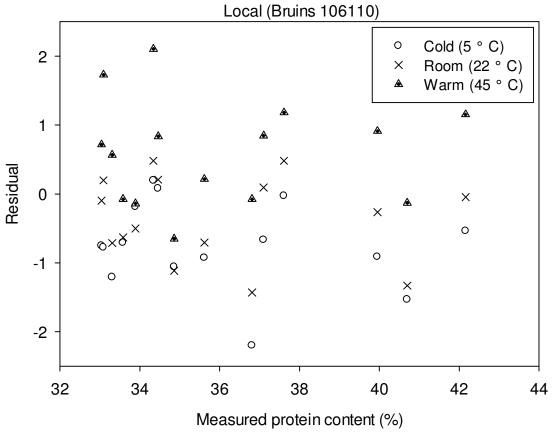
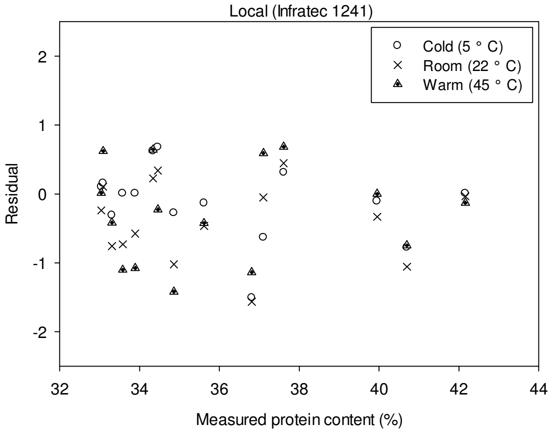
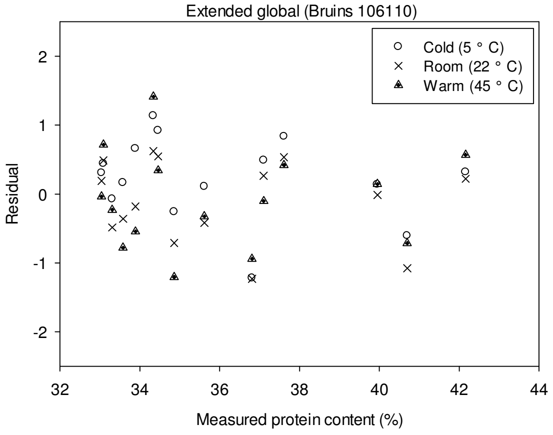
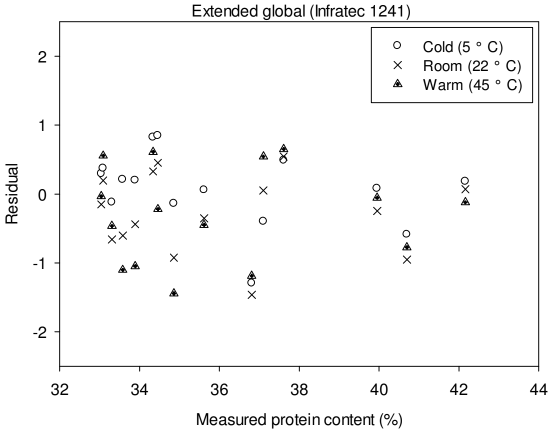


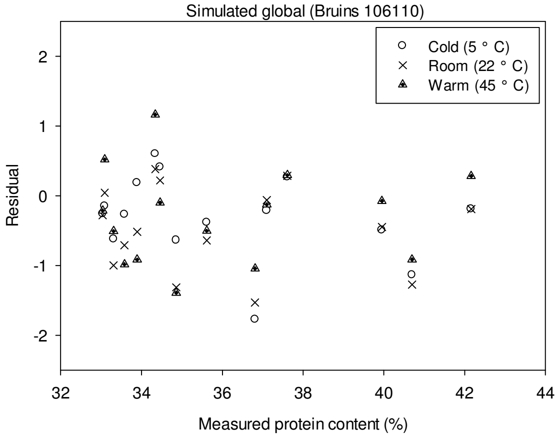
Figure 3. Prediction of the protein content of soybean samples scanned at 5°C (cold) and 45°C (warm) using a model built on samples scanned at 22°C (room temperature) on Bruins OmegAnalyzerG 106110 (a) and Infratec Grain Analyzer 1241 (b)

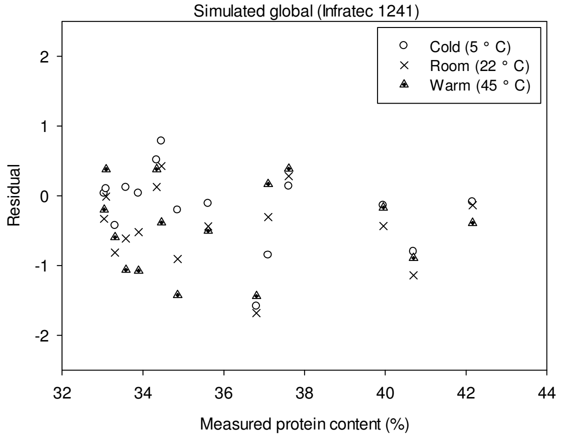

Local (Bruins 106110)


Local (Infratec 1241)


Extended global (Bruins 106110)


Extended global (Infratec 1241)


Simulated global (Bruins 106110)


Simulated global (Infratec 1241)

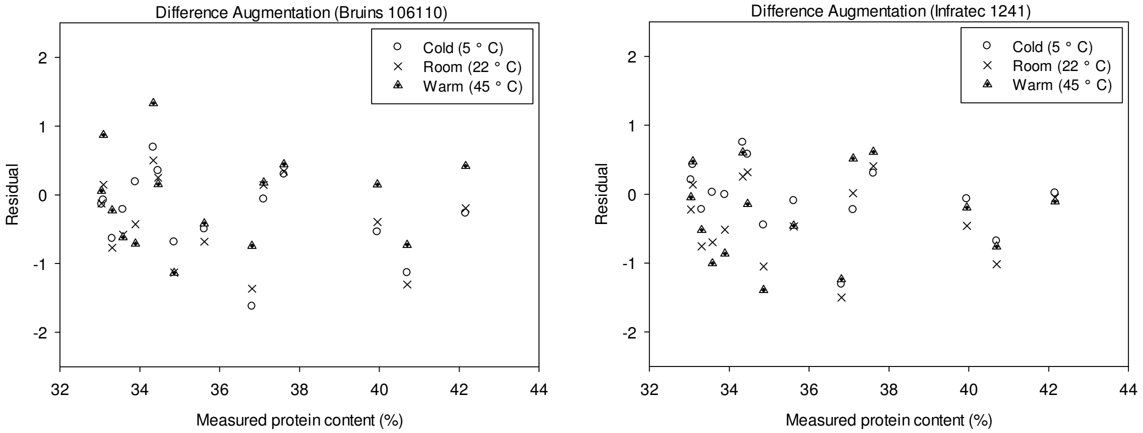


Figure 4. Temperature set prediction residuals grouped by samples temperature vs. protein content for Bruins OmegAnalyzerG 106110 (left) and Infratec Grain Analyzer 1241 (right) with different temperature compensation methods

CHAPTER 5. GENERAL CONCLUSION

General Review of Conclusions

In this study, the possibility of calibration optimization of moisture, protein and oil models was evaluated in terms of calibration sample selection, calibration transfer and temperature compensation.

Calibration model based on a representative sample set is able to obtain similar or even better performance on prediction. The sample selection procedure achieved choosing representative calibration set that covers the variations over the population. The sample selection methods showed that adequate NIR calibration can be obtained based on few representative samples ($n \approx 100$) by significantly reducing more than 98% of the original moisture data set and 90% of the protein and oil calibration set. A comparison among uniform random, Kennard-Stone and D-optimal methods gave approximately similar results with slightly better SEPs for uniform random method. The optimal calibration models were reached with SEPs of 0.14% (n=630), 0.53% (n=429) and 0.54% (n=388) for moisture, protein and oil models, respectively. These optimal models showed better performance compared to the benchmark models built on the entire dataset. This indicates that sample selection procedure not only allows reductions on the initially large set of samples and the complexity of calibration model but also select the subset that represents the initial calibration set well and improves the robustness of calibration.

The transferability of model built on calibration subset could be enhanced by a simple slope and bias correction. Calibration models based on small representative dataset were transferable across units within brand. Models built on master instruments outperformed the local model built on the secondary instruments. Post regression slope and bias correction provides a simple and effective way for calibration transfer for both brands. Calibration transfer based on fewer samples in calibration set cuts down on the number of measurements needed on multiple instruments. Using as few as 20 transfer samples, predictive ability of the calibration model was maintained or improved across multiple instruments and major instrument maintenance.

Calibration robustness could be improved with the inclusion of external perturbations, such as temperature. Near infrared transmittance spectra of soybeans were affected by sample temperature in a non-linear way, mainly due to the absorption changes of the OH-overtones of water. Different temperature compensation techniques were applied to examine the temperature dependence of calibration models for the protein and oil contents of soybeans. Two approaches were found to work well: extended global model with the inclusion of five representative samples scanned at different temperatures in addition to spectra scanned at room temperature; and difference augmentation method that added the generated difference temperature spectra as a noise simulation to the original spectra. The temperature compensation methods did not increase the complexity of model significantly. It is suggested that both the calibration set and the incorporated temperature sample set should cover a representative variation with respect to spectral and chemical information.

In summary, the general conclusions from this research address several aspects of calibration optimization in near infrared spectroscopy. The feasibility of building calibration models with few representative samples has been demonstrated. This provides a means for transferring robust calibration across instrument units and maintaining long-term database, which enhances the efficiency of NIR calibration efforts and saves cost on reference analysis. The predictive ability and robustness of calibration model have been improved by modeling few selected perturbed samples into the selected calibration sets.

Recommendations for Future Work

This research was the first to examine the optimal number of calibration samples to assure a robust NIR calibration model in determining the chemical compositions of grain products. We regard the initial findings as the foundation to attain a reliable useful calibration that can be put to use in industry, which provides a procedure for sample selection. It is useful information to identify calibration samples for people working at industrial-scale calibrations.

Several aspects can be addressed on further research include: 1) the development of an engineering economic analysis helps estimate the life cycle cost of NIR calibration model; 2) the identification of the best samples in advance by the spectra only. Calibration model with analytical accuracy and efficiency is the key to the application of near infrared technology.