Calibration the ISI way

John S. Shenk and Mark O. Westerhaus
Agricultural Sciences and Industries Building, The Pennsylavania State University, University Park, PA 16802, USA.

Introduction

Infrasoft International (ISI) software has evolved through 18 years of research, development and experience. As our understanding of the capabilities of near infrared (NIR) spectroscopy increased, the level of sophistication and user-friendliness of the programs increased. Much of the development has been to improve the calibration capabilities of the software. Calibration is the process of deriving spectro-chemical models that relate the spectra of samples to their laboratory reference values. This paper was written for the food and agriculture industries where calibrations are needed that work now and can be expanded as new samples of a product appear in the future. Our calibration goal has three parts. First, calibrations must accurately model the relationship between spectra and reference values for all reasonable samples of a product. Second, the calibration must provide spectra-based safety tests to assure that unknown samples are predicted as accurately as possible during routine analysis, and third, the calibration must be transportable across instruments.[1]

The spectro-chemical model

A spectro-chemical model relates spectral information to reference method values for a group of samples. This model can be quite complex. Spectra from 400 to 2500 nm are highly variable, containing information on color, provided by electronic vibrations, and on chemical composition, provided by overtones and combination bands of the stretching and bending vibrations from the major OH, CH and NH groups. Spectra also contain the tails of mid-infrared absorption peaks and are affected by surface reflectance, through transmission, scatter and pathlength. Reference methods do not always measure discrete chemical components in the samples. Many reference methods measure different chemical components in different samples.

We have found that the spectro-chemical model is simpler when the group of samples has similar spectral characteristics. But often there is no distinct separation of one group of spectra from another. The easiest way to group spectra is to separate them into the traditional food and agriculture commodity groups, such as wheat, maize, barley, meatmeal, forage and silage. We refer to these groups as products.[2]

The list of the parameters that can be predicted by NIR is increasing each day. But the original list of water, fat or oil and protein is still among the most accurate. This is not because NIR measures these constituents directly, but because this region of the electromagnetic spectrum consists primarily of absorption information about OH (water), CH (fat or oil) and NH (protein). Many other constituents have been predicted by NIR, including fiber, cellulose, amino acids, digestibility and some minerals. Many sample attributes have also been predicted by NIR, including the color of forage and baking quality of wheat. The only way to determine if NIR can measure a particular constituent or attribute is to collect a group of samples, scan them, select the right samples, obtain the most accurate and repeatable reference values possible, perform the

From Near Infrared Spectroscopy: The Future Waves

calibration and test it. If the accuracy is acceptable, the ease of sample preparation and speed of analysis make NIR an ideal analytical technique.

Identifying the right samples

There are many different ways of classifying calibrations. We divide calibrations into four different categories: (i) global calibrations designed to analyze all reasonable samples of a product; (ii) local calibrations for a group of similar samples; (iii) expandable versions of global and local calibrations and (iv) single sample calibration. Each of these calibration types can be developed from a product library. A product library is a collection of sample spectra and reference analyses that represent a product. A product library should represent all reasonable samples of the product, without unnecessary duplication of similar samples. Global calibrations are made from all library samples. Local calibrations are made from a small group of samples in the library that are similar to a new group of samples. Calibration expansion is the process of adding a few local samples to a global or local calibration and recalibrating to improve prediction accuracy. Single sample calibration is the ultimate local calibration.

The four programs CENTER, SELECT, MATCH and SYMMETRY were developed to find the right samples for a product library. These programs rely on converting spectra to scores and measuring the distance between sample scores using Mahalanobis distance. We call this distance measurement the *H* statistic. The first program, CENTER, develops a score generation file (SGF) and uses it to compute the scores and measure the Mahalanobis distance from each sample to the average of the product library. Samples with large Mahalanobis values are flagged as potential mistakes. The SGF currently used by CENTER is principal components, but other methods are being explored. SELECT measures the distance from each sample to every other sample. We use these distances to define neighborhoods in score space. The objective of SELECT is to identify one sample for every neighborhood. Initially, SELECT will remove redundant samples from the product library. Later, SELECT can be used to screen new samples for ones to represent empty neighborhoods. This tends to expand the product library. The MATCH program is used to develop local calibrations by identifying samples that match a specific set of similar samples. SYMMETRY provides three-dimensional displays of the sample scores in a product library. It is useful for identifying and understanding patterns within the sample distributions.

These four programs can help identify the best samples to include in the product library, but they cannot tell which samples are missing. The calibrator is responsible for collecting samples that represent all forms of variation expected during routine analysis. This may include different years or locations, or different formulations when analyzing finished products. Identifying samples that represent all sources of variation results in robust global calibrations. By limiting the number of samples per neighborhood to one, the expense of obtaining laboratory reference values is reduced. The importance of selecting the right samples cannot be over-emphasized. Without the right samples, the true spectro-chemical model will not be found.

Equally important is the task of obtaining accurate reference values for these samples. The software can help to find the right samples but there is no substitute for accurate reference values. Every effort must be made to ensure that the sample scanned by the instrument is the same sample analyzed by the reference laboratory. Moisture weights should be measured immediately after scanning. This is especially true of samples with high or low moisture content. It is important to know your laboratory error. Always analyze some samples as blind duplicates and calculate the standard error of the reference method.

Developing the calibration equation

Two inputs are used by the calibration program. The first input is the calibration file with the right spectra and accurate reference values. The second input is the repeatability file. This file is used to minimize unwanted variation not represented in the calibration set. Examples of such unwanted variation would include any remaining differences among instruments after standardization and variation in sample temperature. The calibration outputs include the equation file containing the intercept and coefficients for absorption values at each wavelength, a score generation file (SGF) used to compress a spectrum into a few scores for the global *H* safety test, and a file of calibration sample scores used to calculate the neighborhood *H* for each sample. Many options must be considered when developing the optimum calibration for a product. Scatter correction can be selected to reduce the effect of particle size or pathlength variation. Mathematical treatments (derivatives) can be used to isolate absorption information in the spectra. A subset of the full wavelength range can be selected to exclude spectral regions that do not contain useful information.

There are many established calibration methods. Partial least squares (PLS) regression works well for most food and agriculture applications. PLS uses spectral patterns developed with the reference values. In addition, PLS can be used with cross-validation. Cross-validation separates the samples into groups for prediction. Each prediction group is predicted once, based on a calibration from the remaining groups. Predicted results are summarized as the standard error of cross validation (*SECV*). *SECV* is used to determine the optimum number of PLS factors that can be supported in the prediction model. Once the optimum number of factors is obtained, all samples are fit to obtain the final equation with the optimum number of factors obtained from cross-validation.

Interpretation of the calibration statistics

Interpretation of the calibration statistics can be confusing. Two basic statistics are used to estimate the accuracy of a calibration equation. The first is the *SECV* and the second is the coefficient of determination (R^2). The standard error of calibration (*SEC*) is not a good predictor of accuracy because it only tells how well the reference values are fit by the calculated regression line. The standard error of prediction (*SEP*) of randomly selected samples can be quite variable, depending on laboratory errors and whether the prediction samples are represented in the calibration set. Our studies have shown that the *SECV* is similar to the average *SEP* from 10 randomly chosen prediction sets. *SECV* is the best single estimate of the prediction capability of the equation.

There are many ways to establish the acceptable level of accuracy for a calibration. One criterion we use is that the calibration error should be comparable to the sampling error of the product. For example, if repeated sampling from a load of hay shows a sampling error of 1 percent, the *SECV* of the calibration should be 1 percent or less. If sampling error is larger than calibration error, the emphasis should be on reducing the sampling error.

The second statistic is R^2. An R^2 value greater than 0.90 indicates excellent quantitative informations. An R^2 value between 0.50 and 0.89 indicates good quantitative information. An R^2 value between 0.50 and 0.69 indicates good separation of samples into high, medium and low groups. An R^2 value between 0.30 and 0.49 indicates good separation of samples into high and low groups and an R^2 value less than 0.29 is not much better than guessing.

In addition to these statistics, the importance of calibration outliers must be considered. *T* outliers are samples with large differences between their reference values and predicted values. Down-weighting samples with large residuals during regression is one method to minimize the effects of *T* outliers. Eliminating samples with large *T* values improves the calibration error but

may not improve the prediction error of new samples. *H* outliers are samples with spectra very different from the average spectrum. These samples, if included in the calibration, will have large influence on the regression model. If they are valid but extreme samples, they will extend the useful range of the calibration. If they are mistakes, they will make the calibration less accurate. Samples with high *H* values should be eliminated only if they are a mistake. When unsure, eliminate a sample with a high *H* value rather than keep a mistake. *X* outliers are samples with spectra that are not modeled well by the PLS equation. These samples should be examined or rescanned.

Monitoring calibration performance

To monitor the performance of the equation we recommend the procedure outlined in the USDA Handbook. Select 10 samples of the product at random for reference laboratory analysis. The bias between predicted values and reference values should be less than 0.6 times the calibration *SECV*.[3] The standard error of a difference corrected for bias [*SED*(C)] should be less than 1.3 times the calibration *SECV*. If the bias or *SED*(C) fall outside these bounds, do not make slope or bias adjustments. Add these samples to the product library and expand the calibration by recalibration.

Almost all calibrations of food and agriculture products need to be expanded at some time. These samples are identified by their global and neighborhood *H* values during routine analysis. We recommend that samples with global *H* greater than 3.0 or neighborhood *H* greater than 1.0 be saved and analyzed by the reference methods. When 10–20 samples with large *H* values have been identified and analyzed, add them to the product library and recalibrate. Be sure to compute new SGF and score files to update the *H* tests for future samples.

Making the calibration transportable across instruments

Testing the performance of a calibration equation on a network is the final task.[4] For food and agriculture applications, we use the following test. If the pooled standard deviation of the predicted values across instruments is less than or equal to the standard deviation of subsamples drawn from a container of the sample, the network is performing satisfactorily. Standardized instruments are just one part of a successful network. The development of robust prediction equations is also essential for successful operation. If all parts of this process are completed successfully, the wavelengths will be within tolerance across instruments, the photometric response will be adjusted to eliminate bias, and all safety test will function across instruments to make the analysis trustworthy.

Summary

In summary, ISI is continuing to explore new and better calibration techniques. These include single sample calibration, neural networks for nonlinear solutions, and the use of indicator variables. However, until new and better procedures are developed, we recommend the following general guidelines for successful calibrations. Use the best product library available for a product and expand it when necessary. Obtaining a good commercial product library from a reliable source is usually a good investment if it can be expanded with your own samples. Global, local, or single sample calibrations can be developed from the product library. Use a repeatability file in the calibration procedure to minimize unwanted variation. Always use the global and neighborhood *H* statistics to make sure the new samples being analyzed are represented in the product library. Use a standardization file to make sure spectra from the host instrument are similar to those from the master instrument where the calibration was developed.

From Near Infrared Spectroscopy: The Future Waves

References

1. J.S. Shenk and M.O. Westerhaus, *Analysis of Agriculture and Food Products by NearInfrared Reflectance Spectroscopy.* Marketed by NIRSystems Inc., 12101 Tech Road, Silver Spring, MD 20904, PN IS-0120 (1995).
2. J.S. Shenk and M.O. Westerhaus, "The Application of Near Infrared Reflectance Spectroscopy (NIRS) to Forage Analysis", in *Forage Quality Evaluation and Utilization*, Ed by George C. Fahey, Jr. American Society of Agronomy Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, Wisconsin, USA, p. 406–449 (1994).
3. J.S. Shenk, M.O. Westerhaus and S.A. Abrams, "Protocol for NIRS Calibration: Monitoring Analysis Results and Recalibration", in *Near Infrared Reflectance Spectroscopy (NIRS): Analysis of Forage Quality. Agric. Handbook No. 643*, Ed by G.C. Martin, J.S. Shenk and F.E. Barton, II. USDA-ARS. US Government Printing Office, Washington, DC, pp. 104–110 (1989).
4. J.S. Shenk and M.O. Westerhaus, NIRS 3 version 3.10. *Routine Operation, Calibration Development, and Network System Management Manual.* Marketed by NIRSystems Inc., 12101 Tech Road, Silver Spring, MD 20904, PN IS-0119 (1995).