NIRS WHITE PAPER

Near Infrared Spectroscopy
for forage and feed testing

History and utility

Initially described in the literature in 1939, NIRS was first applied to agricultural products in 1968 by Karl Norris and co-workers. They observed that cereal grains exhibited specific absorption bands in the NIR region and suggested that NIR instruments could be used to measure grain protein, oil, and moisture. Research in 1976 demonstrated that absorption of other specific wavelengths was correlated with chemical analysis of forages. John Shenk and his research team utilized a custom designed spectro-computer system in 1977 to provide rapid and accurate analysis of forage quality. Early in 1978, this group developed a portable instrument for use in a mobile van to deliver nutrient analysis of forages directly on-farm and at hay auctions. This evolved into the use of university extension mobile NIR vans in Pennsylvania, Minnesota, Wisconsin, and Illinois. In 1978, the USDA NIRS Forage Network was founded to develop and test computer software to advance the science of NIRS grain and forage testing. By 1983, several commercial companies had begun marketing NIR instruments and software packages for forage and feed analysis.

Application of this scientific technique today allows laboratories and equipment manufacturers to serve the livestock industry by providing rapid, highly reproducible and cost-effective analysis of grain and forage via a non-destructive method requiring minimal sample preparation. Perhaps the greatest contribution of NIR-based analysis is that it reduces the total analytical error (sampling and laboratory) because a larger number of sub-samples or sequential samples can be assayed with a limited analytical budget than is possible using the more expensive wet chemistry approaches. This helps producers and nutritionists detect and better manage the variability in feedstuff nutrient composition.

Overview

Near infrared spectroscopy is a versatile, analytical tool for chemical or nutrient analysis based on the interaction of physical matter with light in the near infrared spectral region (700-2500nm). The reflectance or absorption by test samples is mathematically compared with the spectra of reference (calibration) samples that previously have been assayed by standardized and industry approved wet chemistry or non-NIR methods. Reliable NIRS values must come from carefully selected and prepared reference samples to calculate the relationship of absorbance to concentration (Beer's Law: light absorbance = adjustment factor x path length x concentration).

NIRS Forage and Feed Testing Consortium
Excellence in Forage and Feed Testing for the Farmer
http://nirsconsortium.org

Specialized computer software (chemometrics) use mathematical relationship to combine the NIR spectra and accompanying reference chemistry to generate a NIR "predictive model" used to predict composition of the test samples. In robust NIR calibrations, these predictive models are based on hundreds-to-thousands of identity-preserved samples for which researchers have chemical or physical analyses based on reliable reference methods.

Because it is a comparative analytical method, NIRS is a secondary, or indirect method based on regression against a primary or reference method. Consequently, a NIR prediction can never be more accurate than the primary reference method. Every reference method has limits to its applicability and associated error. These limitations and errors must be understood and quantified to decide what degree of analytical error should be attributed to the NIRS prediction model versus the reference chemistry.

Converting light into analytical results

The electromagnetic spectrum consists of many different types of radiation including X-rays, ultraviolet (200-380nm), visible light (380-780nm), infrared (700-2500nm), and microwaves.

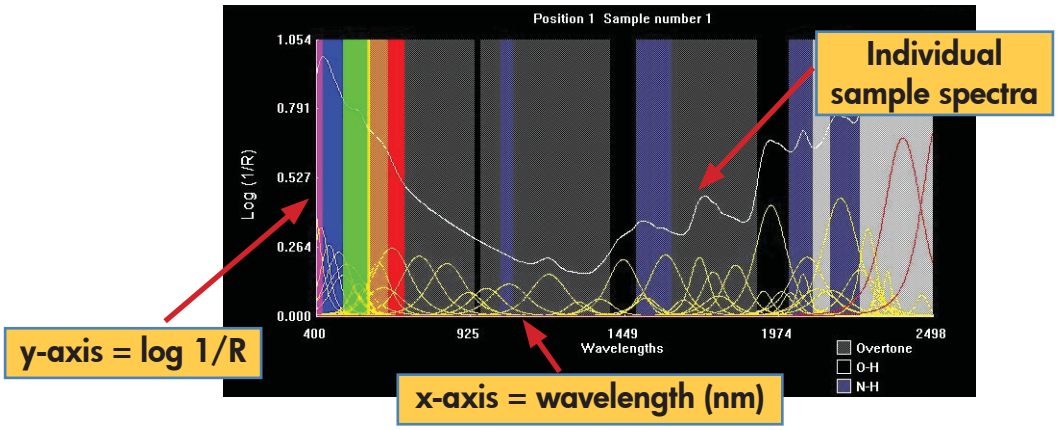
Monochromatic light produced by a NIR instrument can interact with finely ground plant material in a number of ways including as reflection, refraction, absorption and diffraction. Molecules become "excited" and absorb specific amounts of radiation of specific wavelengths. Spectroscopy is possible because molecules react the same way each time they are exposed to the same radiation.

Measurements by NIRS procedures are sensitive to the physical properties of the sample which affect the transmission and reflectance of light. Included among these properties are the physical shape and grind size of the sample particles, independent of whether the means of sample presentation is a ring cup, optical quartz Petri dish, or a natural product cell that uses fresh, undried, unground samples.

In the early 1950's, it became apparent that hydrogen bonding with carbon, nitrogen or oxygen caused stretching, bending, or deformation vibrations responsible for absorption bands in the NIR region. NIR absorption bands are produced when NIR radiation at specific frequencies (wavelengths) resonates at the same frequency as a molecular bond in the test sample. This allows association of a specific wavelength (e.g. 2500nm) with a specific chemical bond vibration (e.g. C-H stretch, C-C and C-O-C stretch) generating a specific spectra that in turn is related to concentration of a specific feed component (e.g. starch). Spectra are clearest for samples with few infrared active bonds and a high purity. More complex molecular structures lead to additional absorption bands and more complex spectra. NIR instruments are much less sensitive in quantifying individual inorganic elements (e.g., calcium, phosphorus or magnesium) or mixtures (e.g., ash) because they are measuring the influence of these "contaminating materials" on the covalent bonds.

In Figure 1, the x-axis displays the wavelengths of light (700-2500nm) being monitored by the instrument. The y-axis displays a typical reflectance instrument response in mathematical terms (log (1/R)). The wavy line represents the spectra ("fingerprint") of an individual sample.

Figure 1. Example NIR Spectra.



Reference analysis and generation of spectra

The individual laboratory or consortium that develops prediction models uses software packages to perform the mathematical calculations necessary to associate the NIR spectra of reference samples with the chemical or nutrient composition of those reference samples. This mathematical process is called "chemometrics." The mathematical equations developed are termed "prediction models," although they are also called "calibrations."

For the highest precision, the prediction model for each trait within each agricultural commodity must be unique. Separate prediction models are required, one for each individual analyte (e.g., dry matter, starch, fat and protein) within specific products (e.g., corn grain, dried corn distillers grains). In addition, modification of a product (e.g., extraction of oil from dried distillers grains) may require development of a new prediction model for a given trait. Although more robust prediction models might be used across a variety of feedstuffs, precision and accuracy will not be as great as prediction models developed with specific and uniform feeds or commodities. Lastly, if prediction models are shared among laboratories, methods for preparing the reference samples for scanning must be fully described and strictly followed by all laboratories.

Because NIR is non-destructive, the calibration sample set is typically scanned with a NIR instrument followed by analysis with the reference methods. Integral to the development of a robust NIRS prediction model is a thorough understanding of the sources of error and minimization of those errors. As a result, it is desirable to have multiple replicates of the sample analyzed by the reference method and also packed and scanned multiple times with the NIR instrument.

The calibration sample set must be developed and assayed adhering strictly to standard laboratory practices. If the calibration set is being developed for a dried, ground sample NIR instrument, then drying conditions must be standardized and the nutrient being measured must be tolerant of the chosen conditions. Specifications employed to prepare the calibration set for analysis must be recorded because those conditions will become the standard method for preparing potentially thousands of samples used in future NIR assays. Sample preparation procedures should not be mixed. For example, NIR spectra for samples

dried in a microwave will differ from that of samples dried in a force-draft oven. Independently, the two drying methods may produce satisfactory calibrations and analyses; but when mixed, incompatibility may introduce errors in the prediction model that cannot be overcome. Because differences in methods for sample preparation are not detected simply by re-scanning the reference samples, new check samples should be assayed frequently by wet chemical analysis to assure that sample preparation methods have not changed from those that were used initially to develop the prediction model.

The reference samples assayed by wet chemistry procedures should be assayed following recommended and standardized analytical procedures reviewed by organizations such as Association of Official Analytical Chemists (AOAC International) or American Association of Cereal Chemists (AACC). The consistency of results among, and within, laboratories from such procedures are characterized by small errors among replicated measurements of the same sample. Further, the reference analyses should be performed in a laboratory that is Grade B or higher in the laboratory performance check sample program of the National Forage Testing Association (NFTA). When a particular analytical methodology may not exist (e.g. for prediction of ethanol yield from corn fermentation), laboratories can develop an entirely new reference method. In general, if the reference laboratory follows AOAC procedures and if the laboratory has a high NFTA performance mark, reference values should be both precise and accurate.

Small particle size for NIR scanning is desirable in order to provide a broad uniform surface for scanning (fewer spectral artifacts) and to minimize variation among the sub-samples selected for scanning. Spectra are quite sensitive to differences in particle size and shape. As a direct result, consistent sample grinding is critical. Compared to grinders that use knives, a grinder that uses abrasion results in differences in particle size and shape. Samples first ground with a knife-type grinder may be passed through an abrasive grinder for final reduction in particle size. With either type of grinder, screens must be in good condition and knives must be sharp. Moisture content of the sample also can alter particle size and should be standardized. One of the largest sources of error in NIR predictions between labs that use the same calibration is a result of differences in sample preparation.

Developing NIR prediction models

One key component of the NIR prediction model is the size and nature of the population of samples covered by the reference method which will be scanned by the NIRS instrument. The sample population should represent the full diversity of plant materials to be scanned. For instance, if the goal is to develop a prediction model for nutrients in corn grain, relevant samples of corn from diverse genetic and environmental backgrounds need to be included in the reference population and assayed by the reference method. If the plant material of interest is only yellow dent corn, Indian maize need not be included in the reference set. Yet, samples of yellow dent corn of diverse germplasm and growing locations/conditions are needed so that the reference sample set fully represents the diversity of the samples to be assayed by NIRS procedures.

To conserve resources, two procedures are used to reduce the number of samples in the diverse reference population. The first is based on the range of values to be measured by

the reference method and the second is based on NIR spectral properties of the population. If the range of reference values in yellow dent corn is 10 to 50 units, then the reference samples included in the set to devise the prediction model must be able to measure that range and the diverse, representative population must span that full 40-unit range. Likewise, NIR spectral properties can be employed to select a diverse reference population. By both procedures, the goal of the sample selection process is to select samples (called the calibration samples or the calibration library) that fully cover all of the expected diversity with minimal redundant information. Extrapolation outside the range of the reference samples is not recommended since accuracy of prediction becomes suspect.

Numerous samples should be scanned by NIR and assayed by wet chemistry procedures to obtain good calibration statistics. A "proof-of-concept" model will utilize 50 to 60 samples. Fully developed prediction models can be built from no fewer than 80-100 samples; however, this number can be greater (1,000's) depending upon the error terms associated with each analyte. The final number of samples required is dependent upon the analytical and spectral diversity within the reference samples selected for developing the prediction model.

The largest source of error in determining the nutritional or chemical composition of a large batch of feed is obtaining a representative sample from the original source. The original source can be a field plot, a silo, a bale, a truckload or a trainload of feed. Samples submitted for analysis typically range from 1 to 5 pounds. Therefore, the process used to collect this sample must not bias its nutrient composition. From a silo of corn silage, a sample that contains either more or fewer kernels than present in the original silage will give analytical results that are biased. From a hay lot, a cored sample that contains either more or fewer leaves than stems represented in the lot will bias results. Similarly, reference samples assayed by wet chemistry procedures must be representative of the samples to be scanned, and the samples being scanned must represent the samples received for scanning. These limitations make sub-sampling and sample splitting critical for all types of chemical analysis.

For developing reliable NIRS prediction models and valid results, laboratories must:

1. minimize sources of error in the entire process;
2. obtain values for reference samples using analytical methods that have high precision and accuracy;
3. standardize sample preparation and analytical procedures;
4. standardize the NIRS instrument;
5. use advanced regression methods like partial least squares (PLS) or artificial neural network (ANN) to obtain accurate, predictive spectral information;
6. perform routine instrument maintenance;
7. analyze only samples representative of the original population;
8. obtain routine diagnostics of all associated instruments and undergo yearly prediction model (calibration) updates.

NIR instruments

NIR instruments can be used to analyze gases, liquids (clear and turbid), slurries, pastes and solids. NIR instruments differ in orientation of the sample to the detector and by the type of detector. By orientation, NIR instruments are of two main types: reflectance and transmittance. For a reflectance NIR instrument, the detector is located on the same side of

the sample as the light source so the NIR light is reflected off of the sample. For transmission instruments, the detector is placed on the opposite side of the sample from the light source so the light is transmitted through the sample (Figure 2).

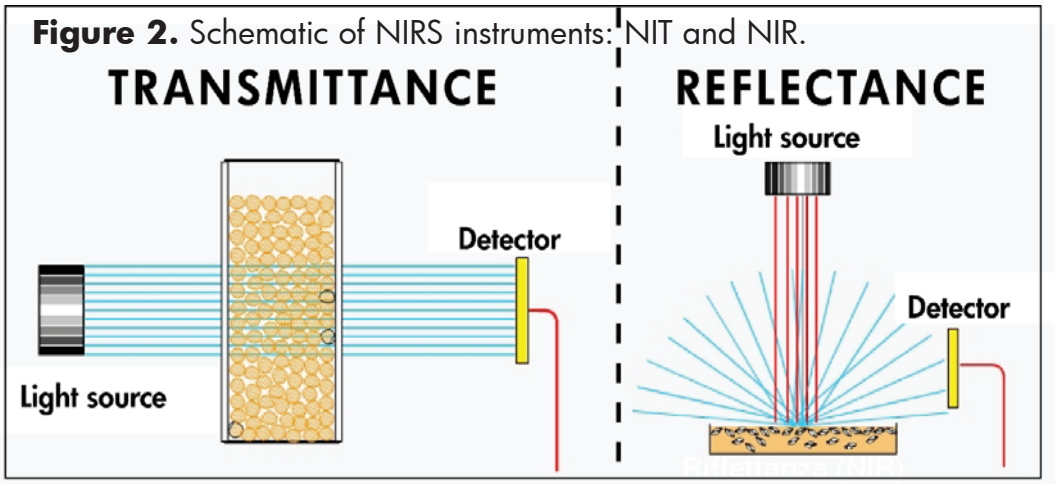


Figure 2. Schematic of NIRS instruments: NIT and NIR.

Detectors consist of two types: monochrometer and diode array. With a monochrometer, a grating or prism reflects the NIR light, one wavelength at a time, from the sample to a sensor. The sensor converts this spectrum into a signal that is processed by mathematical models to produce the analytical result. In a diode array instrument, a bundle of optical fibers sends light to a sensor that is composed of an array of diodes, each tuned to a specific wavelength. Information from each diode is then processed by mathematical models to produce analytical results.

Sample preparation and presentation to the NIR instrument varies widely. Though dried, finely ground samples are often employed, whole grains or fresh, unground samples also can be scanned. Instruments can be stationary in a laboratory or mobile (e.g. on a silage chopper).

Statistics for evaluating NIR performance

NIR spectroscopists often use the term “robustness” or “goodness of fit” when discussing prediction models (calibrations). Error evaluation is paramount to NIR quality assurance and typically utilizes the following statistics:

1. Number of samples in the calibration set (N)

Influenced by the natural variation in the trait of interest. The narrower the range, the more difficult it is to detect differences. Typically 80-100 samples are required for developing an initial calibration with up to multiple-hundreds of samples in a “mature” calibration.

2. Standard error of calibration (SEC)

Defines how well the NIRS prediction model predicts the reference values (calibration sample set) that were used to build the model. Low SEC values are desired. For example, if the reference value is 30 and the SEC is 3, this means 66.7% of the NIR predicted values should fall within the range of 27 to 33.

3. Standard error of prediction (SEP)

Defines how well the NIRS prediction model predicts values for an independent (validation) sample set. Low SEP values are desired.

4. Standard error of cross validation (SECV)

An index of how well the prediction model predicts the reference values in the calibration set when samples are selectively removed from the calibration process. Low SECV values are desired. SECV should closely mirror SEC. Values that differ significantly indicate that the prediction model is weak.

5. Relative prediction deviation (RPD)

The relationship between the SD of the entire population divided by the SEC. This is sometimes referred to as "relative percent difference" in the older literature. High RPD values are desired. For example, if the SD of the calibration population is 9 and the SEC is 3; then RPD = 9/3 = 3. RPD values of 2.0-3.0 allow for adequate screening. Values between 3.0-5.0 allow for improved separation. Values exceeding 5.0 indicate that the prediction model is almost perfect.

6. Regression coefficient (R^2 or RSQ)

The best fit line when predicted values are plotted against the associated reference values. High R^2 values are desired. An R^2 of 1.0 means 100% of the analyte variance is explained by the prediction equation.

7. Standard deviation (or error) of the reference assay (SD)

Determined from replicate analysis of reference samples. Low SD values are desired. The error of the reference method depends upon the chemical method being employed. To characterize reference methods, specific categories (loose, moderate and tight) can be used. Digestibility with an SD of about 2 units is an example of a loose fit. Neutral detergent fiber (NDF) as a percentage of dry matter with a value of 1 to 1.5 is an example of a moderate fit. Crude protein as a percentage of dry matter with an SD of 0.3 to 0.5 is an example of a tight fit. When the reference method is imprecise, the precision of predicting composition of unknown samples also will be imprecise. This also will be reflected as greater NIR SEP and lower R^2 values.

Table 1 illustrates a sliding scale of how "robustness" or "goodness of fit" of an NIRS prediction model varies with the SD of the reference method, with the categories that describe the goodness of fit of the prediction models being favorable, moderately favorable, and unfavorable.

Table 1. Goodness of Fit: NIRS prediction model versus Reference Method.

Goodness of Fit	SEP	R^2	SD reference method
Favorable	SEP ≈ SD	> 0.95	0.3 to 0.5
Moderately Favorable	SEP = 2 * SD	> 0.90	1 to 1.5
Unfavorable	SEP = 3 * SD	< 0.80	2 to 3

An example of a favorable prediction model would be predicting crude protein as a percent of dry matter with a SEP of 0.3 to 0.5 and an R^2 of 0.95. In contrast, an example of an unfavorable prediction model might be NDF as a percentage of dry matter with an SEP of 2 to 3 and an R^2 of 0.80.

The size of the SEP generally varies directly with the SD of the reference method. A reference method must have a low SD if the NIR is expected to provide useful information or be a stand-alone analytical method. Typical standard deviations by laboratories participating in NFTA Check Sample Program are 0.2 to 0.8 for crude protein and 0.6 to 2.3 for NDF.

Users of NIR-predicted values should feel comfortable initiating discussions with their chosen analytical partners with regard to prediction model and wet chemistry statistics. Open communication of P-values and standard errors or confidence intervals by analytical laboratories provides users with deeper insight about analytical precision and accuracy to help avoid confusion and generate greater trust by analysis users.

Summary

NIR analysis as an analytical technique has a long and credible history. NIR is a secondary method that never can be more accurate than the reference method upon which it is based. Statistically robust prediction models allow for a rapid and repeatable assay procedure for nutritional values that help the livestock industry detect and manage variability in composition among and within feedstuffs. The cost-effectiveness of NIR analysis allows the total analytical error (sampling and laboratory) to be reduced because a larger number of sub-samples or sequential samples can be assayed with a limited analytical budget than is possible using the more expensive wet chemistry approaches. To enhance trust, nutritionists, producers and laboratories are encouraged to communicate more fully and openly so that NIR prediction model and wet chemistry statistics are understood more clearly.

References

Approved Methods of the American Association of Cereal Chemists, 10th Ed., Method 39-00.

Berzaghi, P., NIRSC Technical Consultant. Personal communications. Paolo.berzaghi@grainit.it.

Shenk, J.S. 2001. Chapter 16: Application of NIR Spectroscopy to Agricultural Products. Handbook of Near-Infrared Analysis. Editors: D.A.Burns and R.W. Ciurczak. Published by CRC Press. ISBN: 0824705343.

Contributing authors and reviewers

Donald Sapienza – Sapienza Analytica LLC
Paolo Berzaghi – University of Padova, Italy
Neal Martin – U. S. Dairy Forage Research Center, USDA-ARS
David Taysom – Dairyland Laboratories, Inc.
Fred Owens, Bill Mahanna, David Sevenich and Ross Allen - Pioneer, A DuPont Company.

May 2008

NIRS Forage and Feed Testing Consortium
Excellence in Forage and Feed Testing for the Farmer
http://nirsconsortium.org