

Table of Contents

Preface

Introduction

Chapter 1 What is near infrared?

Chapter 2 Advantages and disadvantages of near infrared

Chapter 3 Precautions

Chapter 4 Definitions

Chapter 5 Common sense actions

Chapter 6 First steps

Chapter 7 Feasibility

Chapter 8 Sample selection

Chapter 9 Calibrating

Chapter 10 Additional precautions

Chapter 11 Control standards

Afterword

Copyright © 2012 Mark Electronics, 69 Jamie Court Suffern, NY 10901. This work may not be reproduced by any means, or distributed, without written permission from the copyright owner.

PREFACE

Historically, research in analytical chemistry, or just simply analysis, is mostly dedicated to develop new techniques designed to obtain better and better detection limits and to derive equations that relate to theoretical aspects of various techniques. By so doing, much of everyday usage of techniques has been neglected. Improvements in accuracy and precision of techniques have not been pursued with the rigor they should have been. Further, everyday analyses usually involve many steps, taken by analysts, even with automation becoming more common. These steps have errors in them, which limit the precision of the method being used so naturally accuracy suffers.

The advent of Near-Infrared analysis (abbreviated NIR, or sometimes NIRA, which started in the mid-1970's) in conjunction with chemometrics (a mathematically based approach to extracting maximum information from information-rich sources such as spectra) has revolutionized much of the general area of analyses. As chemometrics is a computer-based system of algorithms (more on this below), no or little analyst error is introduced, the analyses are much more rapid; precision of NIR instruments is almost unprecedented (the only other operation commonly used in chemical analysis that has comparable precision is weighing), the operations are very fast and provide a financial savings due to both the decreased analytical cost and to the faster correction of production variation, whether this is in a plant or in assessing quality of a product.

The technique that has benefited most from the application of chemometrics, to date anyway, is the use of near infrared spectroscopy (NIR). A major reason for this is the spectra are rich in information when compared to, say, visible spectroscopy.

What is amazing about developments and uses in this field is the rapidity and wide extent NIR is being used. A few examples include agriculture, petroleum, pharmaceuticals, biometrics, soil analysis, and so on. Also, almost all nations in the world are using it, some more than others.

This is not to say that NIR is a be-all and end-all. The (relatively) high detection limits associated with the technique prevent its use for trace and ultra trace analyses (although there are steps being taken to drastically lower

the detection limit). Further, many metals and inorganic species in general do not have appreciable absorbances in the near infrared region; this limits the use of NIR for these applications.

We will discuss these points further, below, but here we present a summary of NIR: NIR is a form of molecular spectroscopy. That is, near-infrared radiation is absorbed by molecules (as opposed to individual atoms) via a mechanism that involves the vibrations of the molecules. Put simply, or perhaps simplistically, when the frequency of the light matches the frequency of a suitable molecular vibration, the light can be absorbed. Thus, in situations where a sample can be a clear solution, Beer's law can apply, just as it does for UV, Visible, and mid-IR spectroscopy.

The great power and benefit of modern NIR analysis comes, however, from the fact that NIR can be used successfully in situations where the sample is not a clear solution. It finds its greatest applicability in situations where the sample is a powdered solid, an emulsion, a suspension, or any other type of what would otherwise be considered a "difficult" sample.

The power of NIR, in such cases, comes from two characteristics:

- 1) NIR is only weakly absorbed by most materials, compared to, say, visible or mid-IR radiation.
- 2) It is eminently suited for the application of chemometric analysis of the spectral data produced.

The first point means that NIR radiation impinging on a sample passes through a large amount of that sample. Thus inhomogeneities and other types of flaws in the sample affect the absorbance of NIR radiation less than they would in other spectral regions. In powdered solids, and other discontinuous samples, the discontinuities at the boundaries of the particles in the powder are beneficial rather than detrimental, since they allow spectral measurements to be taken using diffuse reflection rather than transmission. This allows for both reasonable signal levels of the light impinging on an instrument's detector, and reasonable amounts of absorption of that light by the sample.

On the other hand, the strength of reflected light does not follow Beer's law. In fact, no rigorous *ab initio* theory of diffuse reflection exists. However, empirical functions exist that linearize the relationships between the light absorbed and the sample characteristics, and therefore allow the

relationships between the reflected light and the sample characteristics (especially the sample composition) to be determined, via application of chemometric algorithms.

To use NIR in practice, several new skills, techniques and ways of looking at the meaning and application of “chemical analysis” must be learned. The reward for this is the development of methods of chemical analysis that confer previously unattainable advantages in speed, cost, range of applicability, environmental benefits due to non-use of chemicals, and simultaneous multi-component analysis, all from the use of this (relatively) new analytical method. And, to be sure, there are disadvantages, as well, which will be discussed at an appropriate point.

The use of chemometrics has made its biggest contribution, at present, in the near infrared field, but chemometrics can be used with other techniques, too. For example, it can be used in understanding mass spectra. So an understanding of chemometrics that you gain here will help you to understand applications in other techniques.

INTRODUCTION

This overview is written for the new user of near infrared (NIR) chemometric spectrophotometry. It was written by users who have learned the information given below the hard way; *i.e.*, through experience, reading and writing reference works, and some skull work. It is not intended to replace a formal introductory course in NIR. Instead, it is a companion piece and, after some experience has been gained by the user, to be referred to on occasion.

The organization of chapters starts out with an overview of what NIR is. The next chapter discusses advantages and disadvantages of NIR. Chapter 3 covers usages that will prevent damages to the spectrometer and associated parts, and warns against actions that will be generally detrimental to accurate analysis. Chapter 4 deals with definitions and other considerations that should be learned shortly after, or even before, delivery of the system. Other chapters give experimental details, considerations in designing experiments, some pitfalls, mnemonic devices for names, etc (See the table of contents for a more detailed listing).

There is one thing that a new practitioner of NIR analysis must realize about near infrared chemometric analyses; the hardest part of the whole process is what may be called the "hostile environment." This does not refer so much to the physical surroundings as to the acceptance of the technique by others. There will be times that other analysts will not understand the uses or applications of chemometrics. Further, there is much disbelief about the proven capabilities of NIR, so much so that people will not trust what they hear and may want to denigrate your efforts. This will not be done in an obvious way, but when you look for it, you will likely find it. The other side of the coin is to not oversell the technology and make claims that cannot be substantiated. A little pessimism is worthwhile.

All this means that you must be sure to set realistic goals, that are agreed on by everyone in advance. Also, the first NIR project, or even the first few NIR projects, should be for "easy" analyses: those that are known from experience, or from the literature, to give good (*i.e.*, accurate and reliable) results with as little effort as possible. This is always good advice but it appears to have extra meaning for NIR,

since it is necessary to gain the confidence of others about the capabilities of the technology. Analysis of moisture at moderate levels (say, from 0.5 up to several percent), for example, would be a good place to start.

NIR differs from other methods of chemical analysis in several ways. The most basic difference is that the manual manipulations of samples: weighing, dissolving, titrating, etc., are replaced with computerized analysis of the spectral data. In order for this to be successful, the associated computer has to be prepared, or “trained” for the task, so that it can “learn” about the relationships between a sample’s spectrum and its composition. This is accomplished by measuring the spectra of a number of samples of the type to be analyzed, both by the instrument and for the values of the required substance(s) by a reference method (usually the method(s) currently in use for the analysis of the component(s) in question). The numerical data from these measurements are analyzed by sophisticated computer algorithms, collectively called “Chemometrics.” Once the instrument is “trained”, then the instrumental data alone is used to measure future unknown samples, and the reference method is used only to allow QC tests to be made on the instrumental results.

But before starting into the details, one might ask, “What is Chemometrics?” One definition is a “method that uses mathematical procedures to separate the signals of interest from the non-essential ones (doing it mathematically instead of physically) with subsequent measurement of the species or property of interest.” Compare this to classical analytical chemistry where, almost always, there is some physical or chemical separation procedure that isolates the material of interest.

The mathematical procedures are, to a large extent, based on statistical approaches. This does not mean you have to fully understand statistics, but it helps if you are familiar with them.

Chapter 1

What is Near Infrared?

Often abbreviated NIR, Near InfraRed is a type of Spectroscopy. The near-infrared spectral region lies just to the long-wavelength side of the visible region (the part of the spectrum which the unaided human eye is sensitive to) of the spectrum. In the mid-to-late 1960's, Karl Norris, a scientist at the USDA, discovered that the near-infrared region had a convenient combination of useful properties that caused it to be uniquely suited to being used for various applications to chemical analysis, and that could replace methods previously used.

The use of NIR for doing chemical analysis came with a number of advantages:

- 1) Capable of providing rapid analysis
- 2) Capable of providing accurate analysis, often limited by the accuracy of the method used to provide values to compare the NIR spectra with.
- 3) Capable of providing nondestructive analysis. After being analyzed the samples could still be used for other purposes, rather than being discarded.
- 4) Avoids the need for (often harsh, toxic, corrosive and expensive) chemicals
- 5) The concomitant avoidance of the need to dispose of those chemicals.
- 6) Capable of analyzing large and inhomogeneous samples
- 7) The cost-per-test is often much lower than for other analytical methods, especially when large numbers of samples need to be analyzed.
- 8) Multiple components of each sample can be determined from a single measurement of the sample's spectrum; this also keeps the cost-per-sample low

Unavoidably, there are also some disadvantages:

- 1) A moderately large up-front cost for the instrumentation needed, bench space, and sometimes sample preparation equipment.
- 2) The need to calibrate the instrument for each component of each type of sample to be analyzed. In recent times this has been mitigated by the ability of the instrument manufacturers to provide calibrations for common components in universally available samples (agricultural products, for example), but in general, acquisition of an instrument comes with the requirement to calibrate it. The calibration process can be resource-intensive.

The capability of providing rapid and nondestructive analysis, combined with avoiding the need for (often harsh, toxic, corrosive and expensive) chemicals and the concomitant avoidance of the need to dispose of those chemicals, gave a big boost to the then-new technology.

The modern application of NIR uses it to measure the composition of unknown samples (*i.e.*, to do chemical analysis), using techniques invented by scientists at the USDA, techniques which have since been extended by scientists all over the world. It has become a very popular technique in a wide variety of industries due to its speed, accuracy, wide applicability and avoidance of extraneous chemicals. It is widely used in the agricultural, chemical, pharmaceutical, textile and many other industries. The avoidance of the need for extraneous chemicals provides allows a cost savings not only in not having to obtain them initially, but also in the avoidance of disposal problems.

However, there is no free lunch. In order to gain the benefits of NIR analysis, it is necessary to perform a fairly complicated and intricate “training,” or “calibration” procedure. This training procedure involves the measurement of the spectra of a good number of (properly selected) samples, followed by complex calculations that allow the computer program to determine the relationships between the spectra of the samples and their compositions. The calculations allow the computer to create a “calibration” or “model” of the samples, so that future samples only need have their spectra measured, and then the computer can apply the model to those

spectral measurements and determine the compositions of those samples.

The requirement for the training procedure means that NIR analysis cannot be done casually. The training procedure is time-, resource-, and money-intensive. Therefore, in order to be a viable method to apply to any particular analysis, there must be enough “payback” in the form of time and money savings during routine use of the technology, to make up for the time and effort expended in producing the calibration model. However, hundreds (or maybe thousands) of companies, in dozens of fields of endeavor, have found that NIR indeed saves enough money in the long run, to make the use of the technology worthwhile.

The basic operation of “calibrating” or “training” a spectrometer to do NIR analysis is the presentation of a set of samples to an NIR spectrometer, and to have those same samples analyzed by a reference method. The spectrometer measures the spectra of the samples, and then the associated computer program applies sophisticated mathematical calculations to relate those measurements to known properties (including the composition) of the samples. This is called calibrating the instrument, and is a required step in preparing the instrument to do quantitative analysis. We will discuss below, in further detail, the requirements, and the procedures for dealing with them. For now, however, it suffices to note that good results, (i.e., accurate analyses) at the end require two items:

- 1) Good calibration samples at the beginning of the process.
- 2) For each sample, the value(s) for the analyte(s) of interest, as measured by another, reference (often wet chemical) method

Another common use applies a different type of calibration methodology to allow the instrument to automatically identify one or more components of an unknown sample in a given material by comparing the measurements to those made on known materials. This is a form of qualitative analysis.

In both types of calibration, the necessary calculations are performed by a pre-programmed computer, minimizing the amount of information the user has to supply.

Compare this procedure to the classical approach of analytical chemistry. The usual approach is to treat the sample with physical or chemical steps to either isolate the material of interest or to form a known matrix that will not interfere with the subsequent measurement of the material of interest. These initial steps can be time-consuming and are prone to operator error, even if the final analysis is automatically done with instruments.

Having measured the spectra of all the samples to be used to calibrate the instrument, and the values (usually concentrations) of the components (analytes) of interest by the reference method for each sample, there is the need to relate the two sets of quantities, so that future measurements of the spectra of similar samples will allow the computer to calculate the composition. This part of the analytical method relies on a set of computer programs that implement algorithms described by the mathematical discipline called “chemometrics”. The availability of advanced software means that one needn’t be a professional chemometrician in order to calibrate an NIR spectrometer. Nevertheless some understanding of these algorithms is useful, so that when performing the calibration calculations the analyst will be able to direct the computer to produce the most accurate and robust calibration that the data are capable of producing. It is also important to have some understanding of what the computer is doing in order to at least recognize when the computer is producing poor-quality calibrations, even if you don’t immediately know how to fix this situation.

An advantage of NIR is that there often are no, or a minimum of, sample preparation operations needed before the measurements step, the obtaining of the spectrum. Why is this of advantage? At the least, it is much faster, thereby lowering costs. More cost savings are achieved, as hinted at above, by virtue of the fact that no extraneous chemicals need to be bought and stored, and also by virtue of the fact that those same chemicals don’t need to be (properly) disposed of. Anyone who is environmentally conscious will appreciate the benefit of the inherent avoidance of discharges of chemicals into the environment enabled by the use of NIR. Also the minimization of prior steps improves the precision of NIR since it is not limited by the errors introduced by each of these prior steps. The speed of the technique, coupled with the fact that the hardware is inherently capable of being connected directly to many sorts of process equipment leads NIR to be a very useful technique when it is desirable to analyze on-stream, *i.e.*, automatic analysis

with no operator involvement and subsequent rapidity, often taking less than a minute and sometimes less than a second.

The process of performing the calibrations imposes some requirements on the samples that provide the information needed for the calibration process. These can be briefly categorized as follows:

- 1) The nature of the samples
- 2) The range of compositions
- 3) Characteristics needed

In this brief space we can't discuss these sample properties fully. We will, however, describe some of the more important factors, and perhaps in the future we will be able to expand on these discussions.

- 1) **NATURE OF THE SAMPLES:** the samples used for the calibration must be of the same type of materials that the instrument is to be used for in analyzing the future samples, during routine analysis of those samples using the NIR method. This seems obvious, but it includes the warning, for example, that care must be taken to avoid accidental inclusion of different, but similar, materials in the calibration sample set. In the early days of NIR analysis, for example, barley samples would sometimes be accidentally included in sets of wheat samples that were used for calibrating the instrument; this resulted in poor calibrations that were not as accurate for the wheat components as they might have been.

The samples must actually be the same type as the ones to be measured. A pitfall sometimes encountered is to attempt to make synthetic samples, that are "supposed to be like" the real samples to be measured. These attempts usually fail, most commonly because it is not possible in the laboratory, or even in a pilot plant, to recreate the actual conditions that are imposed on the product during actual manufacture. Except in very rare cases, actual process samples must be obtained and analyzed, in order to develop a reliable, accurate and robust NIR analytical method.

- 2) **RANGE OF COMPOSITIONS:** The range of the values of the analyte in the calibration sample set must equal or exceed the ranges of those analytes to be encountered in the future, when the analyzer is put to routine use. It is generally not recommended to try to extrapolate the NIR analysis beyond the range of values of the analytes available when the calibration was performed.

Since it is not always possible to obtain samples that cover the full range of needed values at the time the calibration samples are being measured, one approach often used is to measure the samples that are available, and then add more samples to the calibration set at a later date. While samples might decay or otherwise change over time, the data do not decay, so that it is often possible to extend the range of applicability of a calibration model by adding data to it at a later time, when suitable samples become available, and then recomputing the calibration model.

Another, related, aspect is that it is important to include the expected variations of the other sample components, in the calibration data. Thus, for example, if different components in the samples are present, then they should be introduced into the calibration samples in all combinations. I.e., if components A and B are present, then there should be samples in the calibration sample set that have high and low values of component B along with low values of component A, as well as high and low values of component B along with high values of component A. It is usually not possible to rigorously insure that this condition holds for all possible pairs of components in a sample set, but care in finding and selecting suitable samples can go a long way to improving the calibration set in this respect. There are also some computer programs available that can help select a proper sample set.

- 3) **CHARACTERISTICS NEEDED:** Still another related aspect is that it is also important to include the effects of extraneous variables, such as ambient temperature, humidity, other environmental factors, and variations in process conditions, etc. on the calibration data. Again, it is not always possible to do this

rigorously at the initial data collection, but it usually is possible to collect samples at different times, during different shifts, during winter and summer, etc., and then combining the data sets, as described above, in an attempt to include variations in these extraneous factors.

Chapter 2

Advantages and disadvantages

In Chapter 1 several advantages of using NIR were delineated: faster analyses, little or no operator involvement, lower costs, environmental benefits, improved precision (and potentially improved accuracy) and possible automation. This chapter will delve further into these advantages as well as some of the disadvantages of NIR.

Modern spectrometers nowadays can perform multiple scans and average them in less than a minute and sometimes less than a second. Few other analytical techniques can match this speed. The times needed for calculations performed on the spectra are usually on the order of milliseconds or less. The slow part of the complete analysis in the laboratory (but not in-line) is collecting the sample and presenting it to the spectrometer and the subsequent cleaning of the instrument (or sample holder) to prepare for the next sample. When there is automatic operation the manual preparation and cleaning steps are either not present, or are done by the automated mechanisms.

With little or no operator involvement, the imprecision associated with those steps is absent. Since most spectrometers do multiple scans and average them, the precision is increased. What is the smallest relative standard deviation (RSD)? Typical values are on the order of 0.1% in a developed method. Compare that with volumetric titration, which typically has a RSD of about 0.2 to 0.4%.

Development of an NIR method takes time, especially at the beginning. This may be considered a disadvantage. However, it can be, and often is, at least no worse, and is often less than the time needed to develop the method it is meant to replace. The time savings enjoyed during application of the technology during routine analysis soon makes up for the initial time investment needed to calibrate the system, and these savings keep mounting as the technology continues to be used.

When NIR is used in an automatic mode, especially in-line, the cost advantages are enormous. The major savings gained by use of NIR is not in the lower costs involved in the use of NIR equipment (although those savings don't hurt, either), but in the improvement in the production

capabilities of a process line or manufacturing plant. Continual monitoring of the product drastically reduces the possibility of it becoming out-of-spec (which may then have to be recycled, reworked or disposed of), keeping the amount of expensive constituents to the necessary minimum, etc.; these are the real benefits of NIR analysis. Many types of industries have realized these savings. For example, the petroleum refineries obtain octane ratings in seconds, instead of hours or days. Pharmaceutical firms routinely check 100% of incoming raw ingredients, often analyze every pill or tablet, and use NIR for troubleshooting. Mixing of powders is greatly facilitated by ensuring that they're well mixed. The list could go on and on.

Since fiber optics have their best transmission in the NIR region, samples do not need to be taken to the spectrometer. Instead, the spectrometer can be "taken" to the sample via a fiber optic cable. This has many advantages. The sample may be hazardous and should be contained at least in a hood. The fiber optic cable can be used with the spectrometer well away from the hood and possible associated danger for the analyst. Many spectrometers can be used with multiplexers. In this usage, more than one fiber optic cable is connected to the spectrometer, for example, in a production facility. This allows a single spectrometer to monitor the process stream at several different points. The programming is set up to scan each fiber optic probe at selected intervals. When six cables are used in this multiplexing the cost advantage is enormous.

NIR is called a secondary method in that the samples used for calibration and validation must also be analyzed by another method, which is considered the primary method. This is seen as a disadvantage. It is rarely realized that, except for strictly gravimetric methods of analysis, most chemical laboratory methods are not based on fundamental properties of nature, but are themselves based, one way or another, on comparisons with other methods of analysis. For example, chromatographs are typically calibrated by running some known samples before use; these samples are generally made up gravimetrically, therefore the chromatographic method is a secondary method itself. We find that in general, the reference methods used for the determination of the samples against which we calibrate the NIR analysis are usually secondary methods themselves. Thus the NIR method is actually a tertiary method.

Another supposed disadvantage is the detection limit. This varies but typically ranges down to around 0.1% with a few at smaller values. This is,

in ppm, 1000. The upper limit is 100%. Most other methods of chemical analysis are not capable of analyzing samples where the concentration of the analyte makes the analyte the major component, without dilution or some other way of handling the large quantity of analyte. This is one of the reasons why extraneous sample preparation methods have to be applied to most common analytical methods, but are not needed for NIR.

Chapter 3

Precautions

You want to protect your equipment against damage. This chapter describes some actions that you should not only be aware of but also avoid.

If you are using a “normal” liquid cell, use the same precautions as with, say, UV or visible spectrometry cells. The manufacturer will generally advise you of any special precautions needed, beyond that.

NIR equipment often includes fiber optic probes. These consist of connectors to the spectrometer, glass, or transparent bundles of fibers protected by an outer jacket, usually metal, and the optical end that is immersed in liquids or put close to solids and powders.

The connector from the fiber optic cable should be fairly robust and about the only way to misuse or damage it is to put untoward pressure on it, especially in a sideways direction.

The fiber optic cable part is also susceptible to damage if coiled too tightly or bent around a corner in a sharp angle. These actions can break some of the internal fibers causing at least a loss in signal strength. If this happens after the equipment is put into routine use the analytical results could suffer. If the fiber optic cable is a long one and is used in a laboratory, you may want to suspend it out of the way to keep from accidentally dislodging it. Too long of a fiber optic cable can lead to loss of signal. At the time of this writing, many cables are available that can transmit NIR signals with little or no significant loss of signal over many meters. If you are installing a system in a plant, for example, consult the equipment supplier for the information on usable cable lengths.

Some manufacturers claim that a 1-hour warm-up is sufficient, but we recommend that longer warm-up times be used, except in emergencies. This is because the warm-up period needed until the instrument becomes fully stable is usually a number of hours, 24 or more being the best. How long should the equipment be left on? Roughly, turn the equipment off if it is not to be used for at least four days and a suitable warm-up time is available before restarting. This is somewhat optional as with instruments that use

solid-state devices, as most spectrometers now do, too much temperature change associated with turning them off and on can lead to decreased performance. Some instruments have a provision to turn off only the bulb, or to run it at reduced power, while in “standby” or “sleep” mode. When available, this feature is a good compromise. Consult the manual or manufacturer for exact details. The end of the fiber optic cable is often an interface between the light transmitting fibers and the sample. This interface is a transparent lens usually made out of quartz or sapphire. When used for powder or solid samples, the infrared radiation passes through the lens and interacts with the powder or solid. The light is then collected by another lens placed at an angle. This angle can be 180 degrees from the incoming light or some angle less than that. The angle can be zero degrees, in which case the lens refocuses the light on other optical fibers and is transmitted to the spectrometer.

When the samples are liquids, the end of the probe may be immersed in the sample. A cavity in the probe allows the liquid to enter it. Very often, after passing through the lens and the sample, the light impacts on a mirrored or reflective surface, subsequently reentering the probe back to the spectrometer. This arrangement is called a transreflectance probe, as there is both transmission and reflection. Note that, in this case, the path length of the cavity is twice the distance between the lens and reflecting surface. Typical path lengths range from one to several millimeters. The longer path lengths are used more often for shorter wavelength regions of the spectrum.

There are three substances that can damage the lens. Hydrofluoric acid, hot or boiling caustic solutions, and boiling sulfuric acid are known to attack the lens materials. The other parts of the probe can also be damaged but are not as critical as the lens. If the lens is sealed into the solid part of the probe, care should be taken to avoid immersing the probe in such materials. Consult with the probe supplier for a listing of these.

The cleanliness of the probe and in particular the lens should be inspected often. If the mirrored surface is not shiny, wash it with solvents that are known to dissolve the material into which the probe was immersed. This should be done as a routine both before and after immersing the probe in a sample. Then wash the probe with acetone and dry it with a particle-free air stream.

Wiping the reflecting surface to clean it should be done with caution. Rubbing it too severely can inflict scratches, which could change the reflecting characteristics. If solvents do not remove a fogging of the surface, use a paste made of a non-abrasive detergent and water and gently rub the surface.

One routine practice is to inspect the condition of the probe before taking a spectrum. Even if the probe has not been used for a period of time, dust from the air and other contaminants can accumulate on the critical components.

One way to check the cleanliness of the probe, if your equipment can measure gain, is measure the gain with air in the probe cavity. Then clean the probe. Measure the gain again. The difference in the two gain values should be less than or about 0.005. If the difference is larger, reclean the probe. Continue this procedure until the gain difference is at or below 0.005. If larger values are obtained after repeated washings, obtain a background spectrum and visually examine the spectrum. Compare this spectrum with one you know is good. (For this purpose, a spectrum of the instrument characteristics should be measured when the instrument is new, for reference.) If there are no visual differences, under relatively high magnification, the probability is good the probe is clean.

When using a probe to sample liquids, visually inspect the cavity after immersion. There shouldn't be any bubbles or particles in it. If bubbles are seen, a sharp sideways movement of the probe is usually sufficient to dislodge them. If that doesn't work, remove the probe from the liquid, clean the probe and reinsert it into the liquid while holding the probe at a 45-degree angle. Also, when the liquid starts to fill the probe cavity, stop the immersion until the liquid is at rest. Then slowly complete the immersion while holding the probe at the 45-degree angle. Try to orient the probe so that any bubbles that may exist will rise up out of any internal corners.

One of the parts most subject to a time-related failure is the lamp. For this reason, some instrument manufacturers have provision for turning the lamp off during extended idle periods, while the rest of the instrument is kept on, so it can remain stable (also see comments above, about the bulb in relation to warm-up times).

Some laboratories and plant installations are in extremely dusty areas. If that is your situation, put a high efficiency particle filter (HEPA) on the

incoming air vent. The presence of dust can affect the optics in both the probe and the spectrometer. Electronics in the spectrometer and associated computer can also be affected. Dust on printed circuit boards can lead to peculiar results, especially in humid atmospheres. Not only will the equipment be healthier, you will be also.

The wavelength accuracy should be verified frequently, at least once a month and in some uses, such as pharmaceutical measurements, more often. There are standards used for the wavelength calibration that are composed of metal oxides and hydroxides. Spectra of these standards have very sharp peaks in the absorbance mode. These standards can be obtained from national standards laboratories as well as from commercial sources. Commercial sources often can provide Certificates of Analysis, to certify that their standards are of the highest quality and accuracy.

The first time you do a wavelength calibration, record and post the values near the spectrometer. Compare the wavelengths in subsequent calibrations. Creating a control chart and entering the values from the time when the instrument is new is good practice. The variations should be much less than 1 nm. If they are more than that, your analyses will very likely be in error.

Electrically ground the spectrometer and consider the use of anti-static electricity mats under the spectrometer, computer, and associated equipment. The metal cover of the fiber optic probe should be grounded with a separate ground, to insure it is adequately grounded (the probe can act as an antenna, so it can pick up static electricity very easily). Note that this should not be a separate set of actions; all equipment should be so grounded.

However if the laboratory is short of space, the spectrometer can be even stood on end or any orientation you prefer. The manufacturer may indicate a preferred orientation for the spectrometer, even if only implicitly by where they put the feet on bench space or makes it easier to use.

You should consider the use of a password on the associated computer. One of us had an experience when someone untrained on the proper use of a computer used it without permission. This ultimately required a day's worth of reprogramming it before it could be used again. Also consider attaching anti-theft cables to the computer and spectrometer.

You should frequently back up all your files. The backing up should initially include all the data on your hard drive. Subsequently do incremental backups of the new data generated since the previous backup. Software is available that can do this automatically at preprogrammed intervals, often as an adjunct to hardware backup devices. A good idea is to perform a backup every time you check the wavelength calibration.

Chapter 4

Definitions

There are chemometric terms that may be new to you. This chapter is devoted to defining the most frequently-used terms. This list of definitions is not meant to be exhaustive, either in breadth or in depth. More complete definitions can be found in manuals or reference works, such as the Handbook of Near-Infrared Analysis (Donald Burns and Emil Ciurczak, Marcel Dekker, Inc, 1992, or the second and third editions (Marcel Dekker, Inc, 2002 and CRC Press, 2008 respectively)) and the ASTM definitions (ASTM E131, and ASTM E1655), to say nothing of the extensive discussions in the various books available about Chemometrics. For definitions not listed here, see, for example, the January edition of Analytical Chemistry and a very, very long and comprehensive glossary in the “Handbook of Vibrational Spectroscopy” by Peter Griffiths and John Chalmers, Wiley, 2002. In fact, there are two: one for chemometrics and one for everything else spectroscopic.

There have been many attempts to define many of the terms below. An ASTM subcommittee has been working to achieve agreement among users of NIR.

Macro -- A macro is similar to a computer batch program but is one you develop yourself. They perform a series of computer instructions. The instructions are programmed into the macro by turning on the macro recorder, performing the desired steps, and turning off the macro recorder. When you want to have the steps performed by the computer, go to the starting place in your data or file and start the macro. The capability of creating a macro is generally an adjunct to another computer program, such as MSDOS or EXCEL. Thus the capabilities of any macro depend on the parent program of which it is a part.

Chemometrics -- (Many have defined this term.) A simple definition is the application of mathematics to derive useful chemical information from measurements of chemical interest. Oftentimes these methods require the use of multivariate methods, that is, methods that require more than one data point per sample. Chemometrics often uses signals arising from

chemical interrelations and interactions. This allows more information to be derived from each analysis.

Modeling -- Chemometric calibration demands the development of a model, or calibration, usually by applying one or another mathematical algorithm to the data for which the model is intended to be used. The model is formed by using a computer program. The program correlates spectral characteristics of the calibration samples with the respective reference data to derive the model. The model can then be applied to spectra of unknown samples to yield an analytical result.

Validation – The process of testing a model after it has been developed, for precision of prediction and agreement with other, accepted, values of the analyte, i.e., accuracy.

Prediction -- In chemometrics, prediction refers to the computations that produce the answer or result for one or more constituents or properties. Although this previously was called an analytical result in a classical analysis, the word prediction comes from the science of statistics and more accurately describes the result of a Chemometric calculation.

MLR – Multiple Linear Regression (or multilinear regression) is another modeling approach. MLR is an expansion of a linear regression applied to multiple wavelengths. Simple Linear Regression would use only a single wavelength to create a model that can accurately predict the analyte. The expansion consists of using spectral intensities at many wavelengths, choosing those that have an effect on the relation. For a detailed mathematical explanation, see “The Handbook of Near-Infrared Analysis,” Donald Burns and Emil Ciurczak, Marcel Dekker, Inc, 1992, p. 251. (page 95 and 126 in the 2nd and 3rd editions, respectively)

Often, only one concentration needs to be determined. Only if reference values for more than one component are present are subsequent concentration values calculated.

PLS – This acronym stands for Partial Least Squares. One way to understand this, without using mathematics, is to imagine a data set forms an oblong structure in hyperspace with n dimensions. A partial least squares model is fit to the data to form a straight line along the major axis. The differences between this line and the actual data points are calculated.

These results are called residuals. The partial least squares is again used to “fit” the residuals with another linear relationship, but at right angles to the first. This process is continued. Every time a fit is calculated is called a factor (See Below.).

PCR – This is Principal Component Regression - The definition is best given by the one from The American Society for Testing and Materials (ASTM Vol. 14.01, Standard E131-90). “A mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired degree of accuracy. As successive components are calculated, each component accounts for the maximum possible amount of residual variance in the set of data. In spectroscopy, the data are usually spectra, and the number of components is smaller than or equal to the number of variables or the number of spectra, whichever is less.” Like PLS, it is a matrix-based algorithm. A major difference between PCR and PLS is the underlying equations upon which approach is based.

Factor -- This chemometric term refers to how many times the underlying chemometric modeling equations are applied to calibration data when using, for example, PLS, PCR, Fourier Analysis, etc. A well-defined system is usually sufficiently described using only two or three factors, although it is rare to be able to use so few factors in a real-world calibration situation.

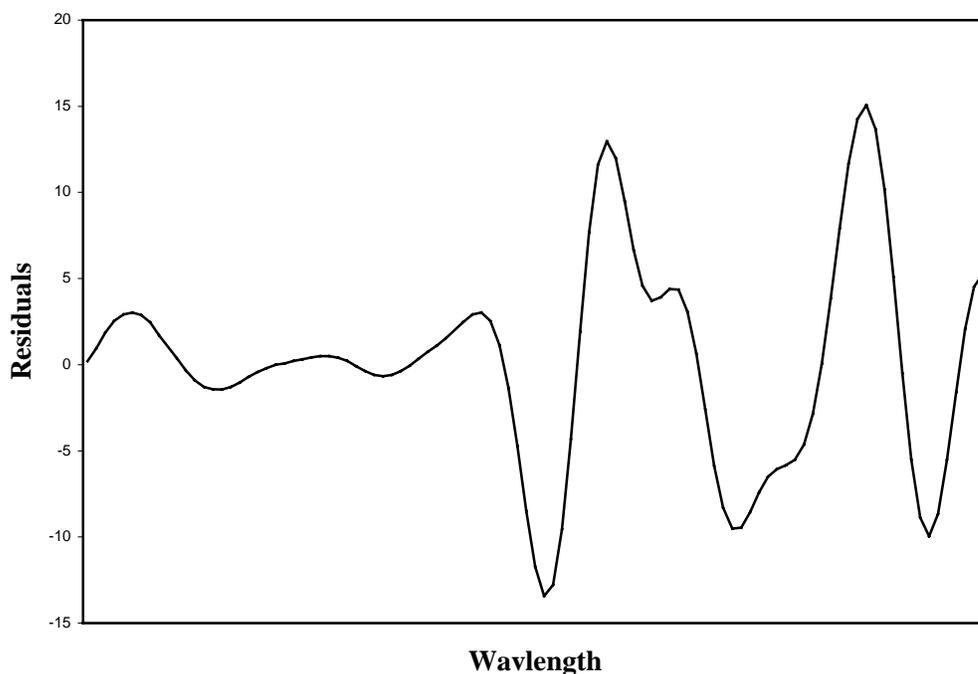
Overfitting – Overfitting is the inclusion of factors (PLS or PCR for example) in the model when those factors model noise that is of the same magnitude as the residuals. In this case, noise begins to contribute to the correlation in a major way. Then, when the developed model is used on spectra of unknowns, spectral noise is included in the calculation. Imprecise and often incorrect predictions are then obtained.

Standard Error of Prediction -- The SEP value is a measure of accuracy that is calculated from how good the predicted value for each sample in a set is compared to the set of reference laboratory values for those samples. The SEP is often close to the standard deviation of the NIR method. If it isn't, there probably is an error in the spectrum or sample. Consult a reference work for this for the mathematical description of how to calculate the SEP.

Standard Error of Calibration -- The SEC value is also a measure of accuracy of the calibration model similar to the SEP except that it is calculated from the same data used to create the calibration model. It is related to the goodness of fit to the calibration or reference data by the chemometric model.

Calibration coefficients – PLS and PCR generate coefficients which are values that are used to multiply the data at every wavelength. See figure 1 for an example of calibration coefficients.

Figure 1. Plot of calibration coefficients vs. wavelength, in nm.



Colthup Chart – A chart that shows the band assignments of the wavelengths of peak absorbances for common chemical bonds.

Mahalanobis Distance – The mathematical location of a sample can be described in terms of a multi-dimensional location in multidimensional space. One measure of distance between two different samples or from a sample to a reference point, is called the Mahalanobis distance.

Mahalanobis distances do not usually have equal equivalent Euclidean

distances in the calibration space, reflecting the differences in properties of the data in different directions.

A guideline for use of the Mahalanobis distance in chemometrics is that when this distance becomes about equal to or larger than 2-3, the prediction is often an extrapolation. However, the distance also depends on the quality of the calibration data, so that after the analysis of several samples, it is possible to determine what the Mahalanobis Distance should be. Maximum acceptable Mahalanobis Distance values can be as large as 3-4. Also, Mahalanobis distances can be used for qualitative analysis. Every substance should have a volume of space centered six or more Mahalanobis distances from all others. If this is obtained, the qualitative analysis should be acceptable.

Chapter 5

Common Sense Actions

Although you don't absolutely need to know computer languages and usages, it is helpful to be familiar with some basics. When you order your spectrometer, software (if ordered separately), and computer, you should ask for the manuals to be shipped immediately, before the hardware arrives. The software manuals will tell you, sometimes in great detail, what you need to know and how to use the computer. That may take a while to assimilate.

You may want to learn matrix mathematics if you don't already have a good grasp of them. If you don't have time to do this, don't worry, as you do not have to know matrices; but it helps to have a basic understanding of matrices when reading literature that contains NIR theory.

File names for routine analyses should be descriptive of the sample set and individual sample. For example, a file name might be wheatnnn, where the "nnn" could be the lot number or other identifying number, but this limits the file to one sample only. One approach to naming sample files is similar to the procedure many companies use to generate lot numbers. The first character is the last digit of the year and the next three are the day of the year. Or if you want to use only three characters, the first is the year digit and the next two are the week of the year. More than one file generated on that date can be distinguished from the earlier ones by letters or numbers in the next position. Yet another naming convention is to use the first character position for the last digit in the year, the next two characters for the month, and the next two for the day. When more than one file is generated in a day, adding a sequential letter allows one to distinguish between files.

Before developing a final chemometric model, you will be generating and testing many trial models. Each trial generates at least one file, and sometimes more than five, depending on the program in use. So the number of files can grow quickly. In this situation, a good practice is to copy the major data files into a working directory and generate the model files there. Once done, the important files can be moved back to the main directory. The name of the new working directory can have the same base name as the main directory with an added suffix of WK to indicate it is a working directory. Once done with the WK subdirectory, it can be purged. If more

than one working directory is used, numerical suffixes can be added to the WK directory name to indicate the order in which they were generated.

Often you will work with a file and know that when you are done, you will purge it some time in the near future. For these situations, generate a subdirectory named MISC (for temporary) or JUNK Copy the desired files into your chosen subdirectory, do the work on them, and copy the ones you want to save to the correct directory. Later, if you wish, you can purge the MISC or JUNK directories without having to worry about accidentally erasing a file you should have kept. Warning: Do not make a directory named TEMP and put it under the C:\ root; a system directory already exists with that name and is used for managing programs.

If you don't want to have the file in a separate directory (such as in MISC), put the letter J (for junk) in the first position of the file name. Then, anytime you see a file beginning with J in your NIR folder, you know you can safely delete it if no other files start with the letter J. Also, because all such files begin with the same character, file manager programs will group them together in a list.

File purging is a task that should be done either at the end of a project or monthly, or some at other regular interval, so that you can pay attention to what you're doing and not make the mistake of deleting a file you still need! In the event you do that, it is possible to recover the file; programs such as Norton Utilities are invaluable for that. However, a warning: in order to be able to recover the file, you must not have written anything else to the disk, before the file is safely recovered. Otherwise, if you write any other data to the disk, the operating system will assign that data to the portion of the disk that was previously used for the now-deleted file, thus overwriting the data saved on the disk.

If you are analyzing samples that may be subject to governmental regulations, you may be required to put some kind of security system on the data, the model, the results or all of the above. At this time, there is no one system of protection that we are aware of that is accepted and specified by the USA government for electronic information and storage. Further, although government agencies in the USA are tending toward having analytical equipment validated, there is no specific, accepted governmental process that we are aware of for validation of NIR spectrometers and software. Anyhow, consult available legal counsel and the appropriate

agency(s). USP, however, has a recommended procedure for how often a calibration should be examined. At the time of this writing, ASTM is working on a draft for just such a need.

Chapter 6

First Steps

There are choices to make when starting a chemometric procedure. Some are critical to first-time successful usage.

- You need to decide which system or problem to start with.
- Ensure you have support from your supervisor.
- Check out schedules of the analysts who will help you in selecting calibration/validation samples as well as those who will be analyzing these samples by a reference method.
- Decide on the deadline for the completion of the development.
- Get agreement on the acceptance levels for the model, *i.e.*, the criteria to use in testing the results of the new method.
- Find out what the range of values for the reference samples should be for the method.
- Agree on how the calibration samples are to be generated.

Neglect any of these points and the project is headed for trouble.

The choice of system, at least for the first few projects you will work on, is not a trivial one. Some concerns that will influence the proper choice are:

- How quickly the method can be developed,
- The importance of the method, how much money/time will be saved, the difficulty of obtaining calibration samples,
- How the method is to be used after it is developed, and
- The general atmosphere around you with respect to acceptance of methods and results generated by a technique that at times appears almost magical.

The choice of initial projects is not easy to make, especially if you are new to chemometrics and NIR. You may focus on more than one project but should not try more than three simultaneously. In this fashion, your chances of at least one success are improved. Although a longer time may be needed to develop up to three methods you will

probably find that there will be intervals in developing each method that can be used for developing other methods resulting in little or no lost time.

How can you select a project that is likely to be done quickly? First of all, the reference method should be precise, samples should be stable, and there should be at least 30 samples covering the entire calibration range as well as variables such as supplier, impurities, etc. (A note on calibration range and reference precision requirements: The ratio of calibration range to the standard deviation of the reference method should be at least 10; the larger the better.) Liquids are usually easier to use than solids. Often, one of the easiest and quickest applications of chemometric NIR is the determination of water at any level down to trace levels.

How can one choose an easy and quick project before gaining experience? Some types of analyses, such as for moisture in the range 0.5 to several percent, are known to “always” be “easy.” Ask advice from consultants, others in similar positions in a non-competing institution, and join a users group, to name a few sources. Also, post questions on NIR list servers and discussion groups. The two you should join are <http://www.idrc-chambersburg.org/cgi-bin/discus/discus> and <http://www.nirpublications.co.uk/cgi-bin/discus/discus> then go to "Bruce Campbell's List" for the latter.

About half the effort in an NIR project, especially when NIR is a new approach, is to convince management not only that the approach is worthwhile but to adequately fund the effort. Part of this aspect of the approach also can be to convince co-workers as well as the management of the area in which the procedure is to be used, such as manufacturing. The remaining half of the total effort is divided roughly in two parts. One is the obtaining of suitable samples and the subsequent analysis by the reference method. The other part is the actual development of the NIR method(s).

If you have a good relationship with your raw material supplier ask if they would send you a large number of retention samples of known and stable composition. By so doing you remove the necessity of gathering and analyzing them yourself. If your supplier asks why you want these samples tell them you are developing a new analytical

method and would be willing to share the results of the new approach. Most suppliers would be more than willing to share parts of their retention samples, especially when you say the method is a non-destructive one and the samples could be returned.

A danger in using retention samples is they may have degraded with time. Ask the supplier if this is a concern of theirs. A note of caution here: some modes of degradation might not affect either the reference analysis or the suitability of the material for its intended use, but could affect the NIR measurement, so try to find out as much as you can about the material. Also make sure the samples will cover the entire range of values. It is even better if there is more than one supplier of these retention samples.

Part of any new project is to ensure there are acceptance criteria and it is in writing. Circulate this and get signed approval. The criteria should be defined quantitatively, be it either statistical agreement with the reference method, speed of analysis, cost savings, improved quality of the product, etc.

Many statistical tools could be used to define how good the NIR method agrees with the reference method. A simple one is the average difference (called the “bias”) between the two. Of course, a large difference is not good agreement but if this exists then the NIR method very likely needs further development. What is a small average difference? The average difference should be less than half the standard deviation of the reference method. For a more exact calculation, consult statistical works to test the equivalency.

It is not unusual for the standard deviation of the NIR method to be less than one-third the standard deviation of the reference method. When this occurs the standard deviation of the differences between the NIR method and reference method are due mostly if not completely to that of the reference method. This holds if the bias between the two methods is small.

Chapter 7

Feasibility

When approaching a new problem to be solved the first question is, "Can this analysis be done by NIR?" This chapter gives you guidelines to help answer the question.

A simple calculation using the standard deviation of the reference method and the range of the concentrations can indicate the degree of success of the NIR method. If the ratio of concentrations divided by their standard deviation is less than six, the development of the NIR method is chancy. This is because a ratio of six or less indicates the reference method results are almost at random. The only way then to have success with NIR is to have a very large number of samples and use replicates in the reference method for each sample.

Since NIR has a lower quantitation limit usually at or above a concentration of 0.3% by weight although sometimes lower in solutions, NIR has a good chance of being utilized. (Note: some NIR users have a slightly higher acceptance level.) If the concentrations are smaller than the lower quantitation limit, one way to develop an NIR method is (1) to find a very strong peak that is peculiar to the analyte, *i.e.*, there are no interferences from other bands in the spectrum, (2) use a very long pathlength, and/or (3) use a combination of spectroscopies that will enhance the signal or suppress the interference signals (fluorescence, lasers, etc). Usually, the difficulty required to analyze at levels lower than determination levels is such that the analysis is better done by another technique.

Visually inspect the spectra of at least two samples, samples where the components are known to have significant differences. If the spectra have noticeable differences, a chemometric NIR method can probably be developed. The spectra should be examined in both absorbance and in derivative modes. (You can see peak differences easier in a derivative mode.) Try different derivative modes also. However, once past the fourth derivative, you are often looking at noise except in the case of major peaks (One of us has used the eighth derivative, but this was only for one or two major peaks). In general,

the complexity of such derivative spectra makes visual comparison difficult.

It is worthwhile to note that when using derivative modes, the peak positions, except for the fourth derivative, are sometimes shifted from where they are assigned in a Colthrup chart. Also, in the first and third derivatives, the band maxima are in a negative direction. (A Colthrup chart is one of if not the most common references you can use in NIR.)

There may not be visible differences in the spectra. This does not mean NIR won't work. This is because differences of less than 1% are not easily discerned but often this variation is usable by NIR. The small differences are spread over one or more bands and the total difference is meaningful to chemometrics. To put it in other words, the difference is integrated over a spread of wavelengths.

The next step, regardless of visible or not visible intensities is to do a minimal experiment. See Chapter 8.

Chapter 8

Sample Selection

It is worth repeating that your analysis is no better than your samples. Proper selection of your calibration and validation samples is critical. There are two sources of samples, “real life” ones analyzed by reference methods or ones made from components. The better choice are the “real life” ones. This is because samples made from components almost always will not contain one or more impurities, may not be mixed appropriately, and depend on knowing the true concentrations of the components. This latter point can consume a large effort at ensuring the known concentration. This effort could be used in analyzing actual samples, thereby avoiding the other possible pitfalls of making samples from standards. On the other hand, “real” samples from a well-running process may not have sufficient variability of the components to create a reliable and robust calibration. One strategy that has been used in such cases is to do both: collect sufficient process samples to get a good representation of the process “signature” on the samples, and augment those samples with synthetic lab samples to increase the variabilities of the various components beyond what the process alone can provide.

If analyzed samples are not available, there is no choice but having to make them from components. If this is done, still try to get several “real life” calibration and validation samples analyzed by a referee method. If only a few “real life” samples are available, then prepare additional samples from components and have a mix of synthetic and real life calibration samples.

Never make calibration samples by sequential dilution of a single stock sample. This may work for a while, but sooner or later an actual sample will be slightly different in some aspect and you will get erroneous results. Even mixing two stock samples to make more samples is sometimes bad, since this is similar to a sequential dilution. You should have at least three such stock samples and mixed in a random ratio. If you don't know how to set that up, consult a statistician on how to do it.

An important part of validation is applying the calibration model to samples whose values are known but are not part of the calibration set. These samples compose the validation set. The predicted values for the validation set must be compared to the "known" values from the reference method. If the average difference is much smaller than the standard deviations of either the reference or the NIR method, and there is no significant slope differences or bias in a calibration plot the NIR method is probably in agreement with the reference method. If the standard deviation of the differences is about equal to the standard deviation of the reference method, the NIR method has a standard deviation significantly smaller than the reference method. This also means the NIR method is in agreement with the reference method.

The validation set should contain samples scanned on two different days with the second day at a much later date, preferably as long as possible after the first time the validation samples are scanned. A comparison of the predictions from the two different dates serves to identify possible sample degradation. If degradation is found, the time for significant degradation can be calculated to insure the analysis is done within an acceptable time period.

A slight timesaving step is to obtain the spectra of the calibration samples and the validation set at the same time, done in a random fashion. Separation of the set of spectra into validation and calibration sets can be done via software. Additionally this avoids a time dependence that may not be apparent.

How many calibration samples should be run? For a preliminary examination, to see if NIR will work, 10 may be enough but 20 are better. There should also be at least 4 validation samples that you can use as a test of the chemometric model.

Each calibration sample should be scanned twice. The statistics are more meaningful when each is run in duplicate. The spectra are not absolutely the same for the same sample for different scans; therefore the calibration becomes more robust when using duplicate spectra scans. Some of the instrumental variations are built into the calibration. Also, for best results, the scanning of the samples should be in random order. If samples are run in the same order as the

concentration change, there could be a time factor introduced due to concentration changes with time.

If there is one set of calibration samples for each analyte, the number of samples should be about 20 times the number of analytes. If the values of the analyte(s) in the calibration samples are known with high accuracy and precision, fewer than 20 calibration samples for each analyte may suffice initially. Also, if the sample matrix is known to be very constant, the number of calibration samples can be reduced. Alternatively, if the reference data are imprecise, the number of calibration samples has to be increased.

However, if every calibration sample is analyzed by the same reference method for each analyte, for example HPLC, the number of samples to use is not 20 times the number of analytes. For example, 30 samples may be enough for two analytes, 40 for three or four, etc. When about 20 analytes are to be determined, the number of calibration samples could be less than 100.

Using only 20 samples, or fewer, is not recommended for a final calibration, but if additional samples are not available, plan to add samples to the calibration base at some later date with the understanding that the initial calibration will give "initial" results and should be used with caution. New calibration samples can always be added to a calibration base. This means having to redo the modeling and possibly recalculating results for samples already scanned. This often goes very quickly. Usually, the new results will not be significantly different from the previous ones, but there is that possibility. Thus, every effort should be made to have a large set of calibration samples at the start so that changes don't need to be made later.

Samples have certain requirements. One is that they are stable. Stability as defined here means that they will not change between when they are taken, scanned with NIR and made or analyzed by the referee method. They should be well sealed. Otherwise they may change composition by evaporation (of moisture, if not other volatile components). Also, when contents undergo temperature changes (air-conditioning, refrigeration, etc) the container will "breathe." This phenomenon occurs when the air space in the container expands due

to temperature increases, pushing vapor out of the container. When the container is then cooled, for example during the night, vapor is pulled in from outside the container, but the vapor also contains water (humidity), carbon dioxide, etc. All this can lead to degradation of the samples.

Calibration and validation samples must have the same matrix and components as the unknown samples. If they don't, the model will not work correctly.

The concentration range of the calibration samples should be larger than the range expected for actual samples. This avoids extrapolation, which is riskier in chemometrics than in most other analytical schemes. If the samples are made by mixing components, developing a broad range is not difficult. When samples are analyzed by a referee method, there is an element of luck in having a broad enough range. One way to better the odds is to have many samples; some of them will probably be close to or at the extremes of the desired range.

If calibration samples are made from only a few components, the concentrations should encompass the extremes and then be randomly scattered along each concentration axis. If there are a large number of calibration samples available, more than you plan to use, select ones that have a roughly even spacing of the analytes over the concentration range.

It is also important that not only should the various components in the samples cover their ranges found in actual production samples, they also need to include all combinations of concentrations, e.g., for two ingredients there should be high-high, high-low, low-high and low-low. It would seem that if more ingredients are involved the number of sample would increase as 2^n . However, experimental designs are known whereby combinations of ingredients can be achieved with far fewer than 2^n samples. These experimental designs are in the category that is loosely called "Statistical Experimental Designs," and are quite effective in reducing the number of samples (and amount of work) needed.

Finally, do not neglect chemical knowledge, experience, and intuition. Although chemometric models can be developed without any knowledge of the chemistry of the system, there are traps that can befall the unwary. One trap is having an impurity in the system that has a concentration/response that is proportional to the component(s) of interest. The spectral characteristics of the impurity will be included in the model. Later, when samples are analyzed, the spectral impurity bands will contribute to the analysis. If the impurity is not present in the sample at the same proportion to the analyte as in the calibration samples, the analysis will be in error.

The gist of this is to know as much about the system as possible including the chemistry, how the reaction occurs, and definitely how the samples were taken. Make certain the samples are representative, the sample containers were clean and that no changes have occurred between taking of the sample and the analyses.

Chapter 9

Calibrating

Once the spectra and the reference value of the calibration samples have been collected, modeling begins. Considerations are how the spectra are to be presented, whether the data should be modeled in transmission, absorption, or a derivative of absorbance, or some other data transformation.

Most NIR models are linear, requiring the response be proportional to the effect, such as absorptivity versus concentration. Therefore, the mode of representation has to be linear. Percent transmission is not linear with concentration so it cannot be used for most models. Even absorbance is not always linear over all concentrations, but it often comes close enough that deviations can be neglected.

COMPONENT CONCENTRATIONS

Previous belief was that the units for analyte concentration, in a calibration environment were all equivalent, in the sense that the calibration would accommodate the differences implicitly, by increasing or decreasing the calibration coefficients to compensate for the difference in numerical magnitudes of concentrations rendered in different units. Recent research has revealed that, contrary to expectations, and to previous belief, not all units of concentration are equivalent that way. A simple demonstration would be to consider a binary mixture, with 50 weight % of each component. Unless the components have equal densities (which would be an unusual situation, in general), the volume percentages will not be 50/50. Similarly, a mixture made up of equal volumes of the two components will not contain 50 weight % of each component. Thus the two measures, weight % and volume %, are not linearly related to each other. Clearly, if they are not linear with respect to each other, then only one of them can be linearly related to the spectroscopic measurements; the other one, of necessity, must be non-linearly related to the spectroscopic measurements. A recent paper (Appl. Spect., 64(9), p.995-1006 (2010)) has demonstrated that, in fact, it is the volume percent that is linear with the spectroscopy. To state it

differently, the spectroscopy is sensitive to the volume percentage of the components in a mixture.

If significant non-linearity is suspected and cannot be corrected by changing the units in which the analyte concentration is measured, the easiest solution is to break the concentration range into two or more parts, based on concentration, then make a calibration model for each part and see if the individual parts have a better fit than one overall model. If the fit is better, there may be a non-linear relationship. If such a non-linearity is found and the two, or more, models have internal consistency as well as meet the requirements then the calibration is in many parts. An additional “overall” calibration can also be developed; to know which range the sample is in, so that the proper reduced-range model is applied. The software associated with the spectrometer should be able to handle this situation.

Alternatively, the property you are trying to model may not be linear even if spectra are converted to absorbance. Then, you may need to know the theoretical equation relating the property to concentration. (Viscosity is such an example. The logarithm of viscosity is usually proportional to concentration. Other, special, cases also exist. An example is water-alcohol mixtures, wherein interactions (hydrogen-bonding) between the ingredients affect their behavior so that non-linear spectral effects are seen.) In some cases an equation can be used to mathematically transform the calibration data into a linear dependence. A common transformation for reflection spectra of solids is the Kubelka-Munk one. However, if the theoretical equation is not known, taking the logarithm of the calibration data may be sufficient to linearize it. In any case, if the calibration data had been mathematically transformed, the results of analyses from the model based on the transformed calibration will have to be re-transformed. This can easily be done by adding a batch, macro, or spreadsheet program that will convert the predicted results into the desired units.

Spectra should be converted into absorbance. There's a very good reason that $\log(1/R)$ is the de facto standard for data collection in the NIR, but it has nothing to do with either optical diffusion, other optical effects or the mathematical processing (well, maybe that last isn't QUITE true).

The reason is this: for the usual reasons of being able to correct for instrumental drift effects, etc., a reference reading of some stable material (often a white ceramic disk is used in the NIR) is normally used to divide the sample reading, to compute the reflectance (R), analogously to the use of a reference beam in the mid-IR (to compute T). This is done for the same reason, also: accommodating and compensating for changes in source brightness, etc. The data is then converted to absorbance ($\log(1/R)$), again analogously to what's done in [mid-IR | transmission] spectroscopy. The absorbance data is then used to construct a calibration model, of the form

$$C = b_0 + b_1A_1 + b_2A_2 + \dots + b_nA_n$$

where C is the computed concentration, A_i is the absorbance at the i^{th} wavelength and b_i is the calibration coefficient for the absorbance i^{th} wavelength. This equation is valid even for PCR and PLS calibrations as well as MLR calibrations, since they can be put into that form.

The problem with this scheme arises when something happens to the reference reflector, e.g., it gets dirty and changes differently at different wavelengths, or even if it changes uniformly across the spectrum. If you use any transform of the spectral data other than $\log(1/R)$ then in principle at least, if not necessarily in practice, you would have to recalibrate your analytical model every time that happens. If you use the $\log(1/R)$ transform for the spectral data, then any spectral variations you might have, as long as they are stable after the change, can be shown to affect only the b_0 term of the calibration model, and thus be compensated for by a "bias correction" (as it is called), which is relatively simple to measure and compute. It is to avoid the need to recalibrate an analytical method for every little glitch that might happen, that is the driver for use of $\log(1/R)$.

And THAT's the reason for the near-universal use of $\log(1/R)$. Everyone working in NIR at the beginning knew this, but apparently the knowledge is becoming lost.

Conversion of the spectra to absorbance is often done automatically by the data collection software. The next step is to decide whether or not to use a system to correct for baseline variation. Few spectrometers yield baselines that are flat enough to use without some correction. One common correction method is to have two spectral points that are known to be at the baseline. A straight line across the

spectrum is generated using these two points. This straight line is then used to correct the baseline at the other points in the spectrum.

A more common procedure to correct for baseline variation is to take derivatives. If the baseline has an offset with no slope, taking the first derivative corrects for this. If the baseline has an offset with a slope, taking the second derivative will correct for baseline variation.

Although you may wish to try these three approaches to baseline correction, for your first few applications you should follow the recommendations of the spectrometer supplier.

If you have extra time, you may try calibrating using all three baseline correction methods mentioned above.

If derivative spectra are used, which one is better: first, second, third, etc? First derivative spectra are usually better than higher derivative ones because any time a derivative is taken, there is some imprecision introduced in calculating it. This imprecision adds "noise" to the spectra and the noise degrades the spectra. Further, taking higher derivatives magnifies the noise that is already present in the spectra. Of course, there are times that taking higher derivatives gives definite advantages. But remember, when a complete spectrum is used for the analysis, often like in PLS and PCR, information about changes in band position, intensities, and interactions are still present in the lower derivative spectra. The model can detect these changes. Taking higher derivatives doesn't usually make any difference to the mathematics. Therefore, Occam's Razor should be used: If a model of a second (or higher) derivative gives equivalent results when compared to results from a model of first derivative spectra, the simpler model is preferred.

Calibration can now begin. There are choices to make. The common chemometric approaches are MLR, PLS and PCR. There are many others but for the first trials at calibration, these are recommended.

As mentioned in the definition of factors, a careful choice of how many factors to use exists for PLS, PCR, Fourier analysis and others that generate factors. How many factors should be used? This depends on how precise the model must be and the quality of the reference

data. A quick rule of thumb is there should not be more than one factor for every ten calibration samples.

When developing a calibration, all or part of the spectra may be used in PLS and PCR. Some wavelength regions may not contribute to the calibration and often can be disregarded in the calibration. One advantage of doing this is avoiding some of the noise and correlations that are essentially meaningless. Another advantage is that some future samples may have impurities that give rise to bands in these omitted regions. This approach is more meaningful when a limited number of calibration and validation samples are used. Analysis time is not a factor as the calculations are performed extremely fast. Also, few spectrometers that scan the wavelengths will scan the omitted regions as well as the included ones so that scanning time is the same.

How does one select regions that can be omitted in the calibration? Of course, in MLR this is not a problem as only a few selected wavelengths are used. PLS and PCR software generate regression coefficients that can be used for selecting wavelength regions. Regression coefficients that are large in regions where no bands are present often indicate a large contribution of noise to the calibration. Theoretically, if there's no information in a spectral region, the coefficients in that region should be small. In these cases those wavelength regions can possibly be safely omitted.

A more exact way to find regions that do not contribute to the calibration is to examine the regression coefficients by overlaying them on spectra. That is, all of the products, resulting from multiplying the regression coefficients by the absorbance at each wavelength, are added together to give the analytical value for the analysis. If, say, absorbance was used in the calibration, positive regression coefficients that match positive bands indicate the chemical bond at that wavelength contributes positively to the calibration. A negative regression coefficient (or positive one) matching position with a positive (or negative) band means the chemical bond corresponding to that band generates a negative portion of the calibration, and is probably correcting for the influence of an interfering material. Of course, when both the regression coefficients and the band intensities are negative, as are seen in derivatives of the absorption, the contributions are

positive. In this case, the calibration model is probably using an absorbance band of a constituent in the samples that is correlated with the analyte.

If the spectrometer is a wavelength scanning one, a choice is to be as to which software should be used, PLS, PCR, MLR, etc. The first three should be used initially. When the results from the validation samples are compared, and to some extent the calibration statistics, a choice of the best one may be apparent for the continuation of the method development.

You will probably find comparisons between results generated by each calibration procedure show equivalence, or near equivalence. Equivalency is good evidence that each model does not incorporate significant amounts of spectral noise, or, overfitting is not of concern. This is a guideline, not a guarantee. To use this comparison, the number of factors in PLS and PCR should be about the same in each model. For example: Develop a model using PLS. Generate another model using PCR on the same calibration set and same number of factors. Test each on the validation set. If they give equivalent accuracy and precision values, overfitting is probably not present when using that number of factors.

If one or more models, when tested on the validation set, give results that fit the acceptance criteria, further modeling may not be necessary; it may even be counterproductive in that you are using time that could be better spent elsewhere. That is, an absolutely correct validation may not be much different. Additionally, an absolutely correct validation should use a very large number of calibration and validation samples. That is, you may not have developed the absolutely best model but pragmatism is important here. Consider the number of possible models that could be developed. This number is in the thousands, if not the tens of thousands.

Several other procedures you should routinely consider using in model development are spectral smoothing, deleting ends of the spectra, and band removal.

Spectral smoothing may give a better model, but again this falls in the realm of performing a mathematical operation on the spectra, which

could distort them and cause trouble with actual samples. This doesn't happen often, but be conscious of the possible danger.

Another commonly used procedure is to delete the ends of the spectra and compare modeling results with models that use the entire spectra. The reason for deleting the upper wavelength part of the spectra (from 2300 or 2350 nm up) is that with a fiber optic probe and many detectors this part of the spectrum has little spectral energy and lots of noise. The signal-to-noise ratio is much worse than lower wavelength portions of the spectrum. The part of the spectrum from about 1200 - 1250, or 1200 - 1300 nm often doesn't have many large bands of interest. Deleting that part of the spectrum often leads to faster and better calibrations. If, however, you are using spectra that incorporate the 800 to 1200 nm region, don't automatically delete that region. The spectral noise is low there and sometimes good models use that region. This can be done with longer path lengths, which enhance the band intensities. Compromise measurement conditions are often needed between different spectral regions where band strengths differ. Note: Some spectrometers scan in the visible region also, so both the visible and near infrared regions can be used.

When developing a model, pay close attention to the plots of the regression results, especially the regression coefficients. These plots resemble spectra. The "regression bands" that are more intense are the ones that have greater influence on the model. If one or more of these bands are in a wavelength region that does not correspond to a band in the sample spectrum, you may have a bad problem with noise, there is a component in the sample you were unaware was there, or the model is using a very small absorption band. In any case, delete that part of the spectra from the sample spectra, remodel, and retest against unknowns. If a significant improvement in results is seen, use the model based on having that part of the spectra deleted.

There are guidelines on how to choose the number of factors. One is to think of the plot of percent of the spectra-described versus the number of factors as a curve with the left end being high, and then decreasing to the right, reaching a minimum and finally rising. (See figure 2 and the expansion of figure 2 in figure 2a.) This plot is from an actual calibration. The plot of predicted values versus the reference

values is given in figure 3 as an indication of the goodness of the calibration.

Figure 2. Percent variation explained by PLS on second derivative.
Note the slight upward slant of the line at the higher factor numbers.

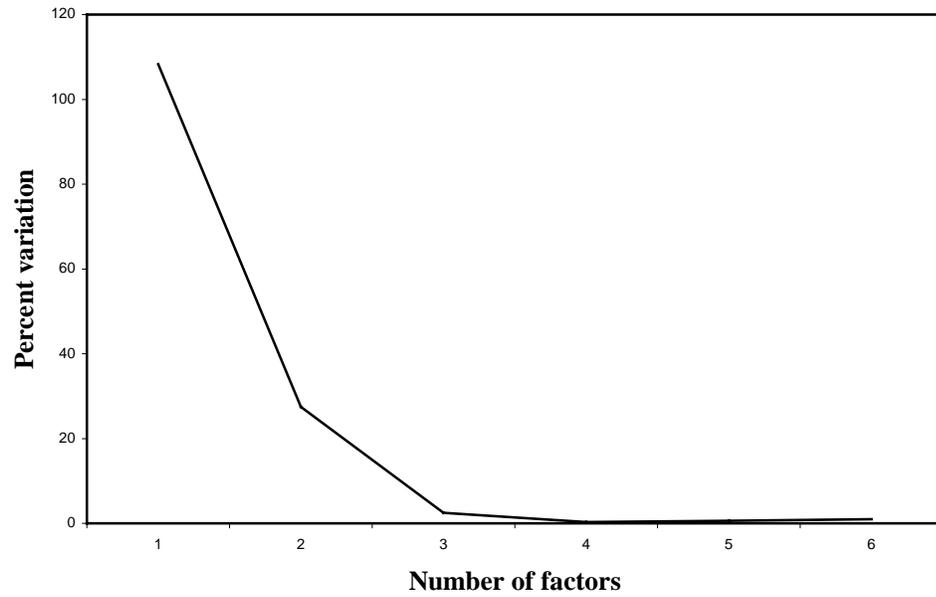


Figure 2a. Expansion of figure 2. The points are at factors 3, 4, 5, and 6.

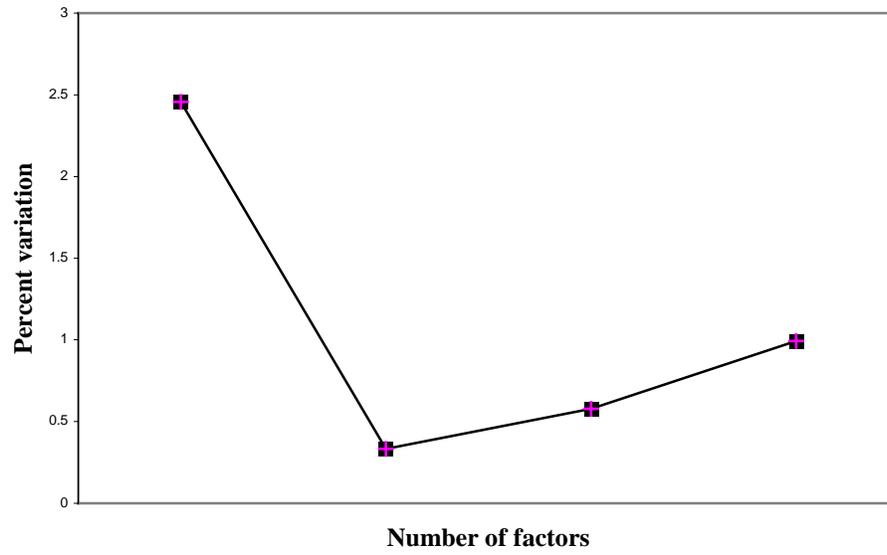
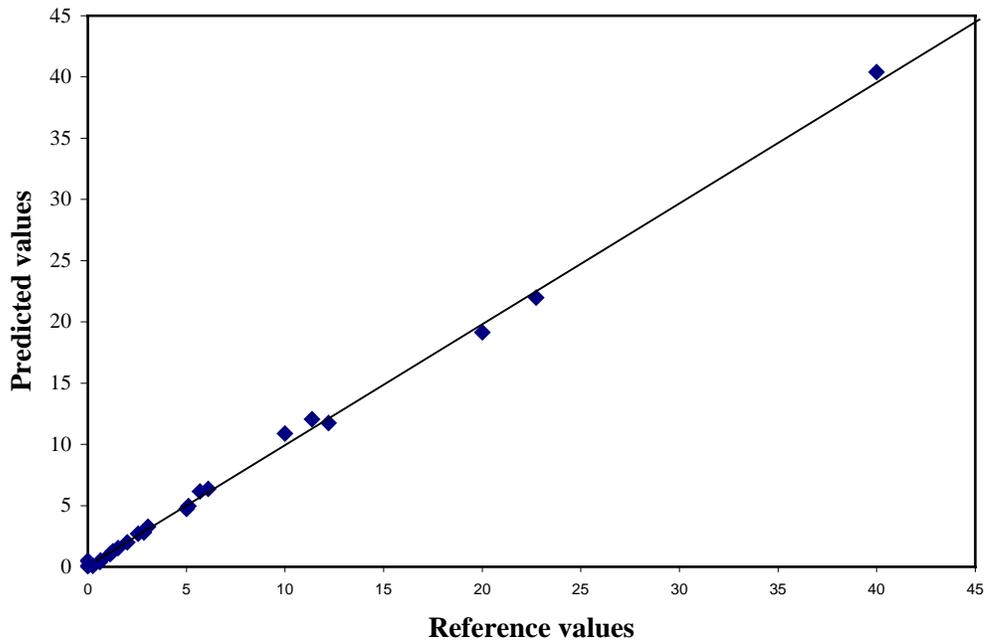


Figure 3. Plot of predicted vs. reference values. A 45 degree line was drawn in.



The rising is very often due to inclusion of noise in the model. The number of factors to use, or at least test initially, is the number of factors corresponding to the inflection point. An even better situation is when the number of factors that are in a very good calibration is to use factors that are less than those at the inflection point. This gives a more robust calibration. One way to examine for fewer numbers of factors, rather than guessing, is to see if the percent of spectra-described that is shown in an associated software plot is fairly flat to the left of the inflection point. If this is so, try calibrations using the number of factors that are at or just to the left of this “flat” region. (Note: Your software may not have the capability to plot the percent of spectra described, but the ability to make such plots is becoming more common.)

Another consideration when choosing the number of factors in developing a model is the chemistry of the component. If the component is involved in a large number of equilibria, then the number of factors increases. For example, five to seven factors are usually needed when developing a model for formaldehyde. This is because formaldehyde participates in many equilibria.

How precise can results be? It is not unusual to achieve replicate results on unknowns with a relative standard deviation of slightly less than 0.1%. Almost all other commonly used analytical techniques have larger relative standard deviations. This means NIR results will usually be more precise than those from the reference method.

Finally, always visually examine the spectra obtained from both the calibration and validation samples. This point cannot be overemphasized. An extreme example happened to one of us. For some reason, perhaps a bubble in the light path, the spectrum of a sample was almost pure noise. Other examples would be a sudden jump, that is, a discontinuity; a spectrum that is much different than the others in terms of noise or positions and intensities of bands for similar samples; etc

Chapter 10

Additional Precautions

Although examples of various pitfalls have been given above, they are summarized here together with a few others. This list is not exhaustive.

The first pitfall is to not have acceptance criteria at the start, or to have acceptance criteria but not get agreement from all involved. Without this you are trying to hit a moving target. The second pitfall is to have an impurity in the calibration samples that is proportional to concentration of the analyte. The model will include the spectral contribution of the impurity and when unknowns don't have the impurity present at the same proportionality, wrong answers will result.

Overfitting is a problem. If results appear peculiar, try another model with fewer factors.

Non-linearity of response can occur. Since the models are based on linear equations, some error may result, as Beer's Law is not absolutely linear plus spectrometer response is not linear when close to its upper absorbance range. If the error is small, it may be acceptable. Otherwise, split the concentration range into two parts and model each part separately. Non-linearity may be due to trying to fit a characteristic that is not linear with spectral responses (such as viscosity). Then, theoretical equations covering the situation should be linearized with respect to concentration.

A subtle non-linearity can arise when mixing materials that have partial molar volumes that are not additive. The best-known case of this is mixing alcohol and water. The volume of the mixture is less than the total volume of the separate components. Therefore, samples should be made by weight rather than by volume. Even then be aware the amount of material in the pathlength will vary. This can be adjusted for, however.

The calibration samples should cover the total range expected for the unknowns. This may not always be possible. In that event, don't

accept results that have more than about 5% extrapolation, and in extreme cases not more than 10% extrapolation.

Pushing detection limits too hard can give poor results. A quick way to see if this is happening is to measure the noise level. The limit of detection is defined as three times that noise level. The quantitation limit is nine times that noise level. So if your lower levels of the measurement of interest are below the quantitation limit, you could be obtaining erroneous predictions

Distorted peaks often result from path lengths of the probe/cell that are too long.

Emulsions can be inhomogeneous and because many emulsions are opaque, you can't see the inhomogeneity. Make sure all emulsions are well mixed. The same goes for powders.

Chapter 11

Control Standards

You should use a control standard, both when performing the calibration and also when doing routine usage. Control standards, analyzed with unknowns, can be either a typical sample, with all the attendant dangers associated with keeping material from changing due to storage, or lack of, conditions, or a different material. The values from the continuing analyses of the control standard will help establish the day-to-day variability of the method. You may want to try two control standards for your first few sets of samples. See which is better (although this may take a few months before possible changes in stability are evident).

Comparing values of control standards is easily done with control charts: Not only can you spot trends in the results of the control standard, but you will be amazed to see how repeatable and reproducible the results are.

Afterword

This overview is not exhaustive, nor was it meant to be. Recommendations for new chapters, additions, and operations are always welcomed.