

Raw Material Identification with FT-NIR



Abstract

FT-NIR raw material identification is a well-proven and widely used technique in the pharmaceutical industry. The purpose of this application note is to provide details pertaining to the technique, the available instrumentation and the sampling options. A brief introduction to the use of PLS as a discrimination algorithm along with the steps required to develop and validate identification models are provided. An example relating to the discrimination between Pseudoephedrine Hydrochloride and Pseudoephedrine Sulfate is used to demonstrate that NIR provides fast and accurate results.

Overview

1. Problem description
2. NIR Solution
3. Operational Aspects
4. Performance Example
5. Summary

1. Problem description

Raw material identification is a tedious and time-consuming task of the quality control (QC) laboratory. Furthermore, European regulations requiring 100% ID testing for each container of raw material received drastically complicates matters for QC laboratories. Labs tend to be short of space, to have limited equipment and to be short of highly skilled lab technicians. The QC lab is, however, still expected to handle the amount number of samples and analyses required by new regulations. This can result in a heavy laboratory backlog that, in turn, may increase raw material inventory and production cycle time.

The new FDA PAT initiative offers the opportunity to build rather than test quality into finished products. The key to building quality into a process is to focus on the relevant relationships between raw materials, manufacturing processes, and environmental variables and their effects on quality. The goal of the PAT framework is to design and develop processes that can consistently ensure meeting predefined quality requirements at the end of the manufacturing process. FT-NIR techniques provide valuable information related to both physical (e.g. particle size, morphic form, etc.) and chemical attributes. This sort of analysis is commonly referred to as “qualification” rather than “identification.” A qualification method entails a meticulous comparison of the sample with the spectral features of those in the calibration set rather than just a spectral match as is performed in an ID analysis. The application of this technique can save a substantial amount of money by minimizing the production of an out-of-specification finished product (and all the huge costs associated with an investigation) by preventing poor quality raw material from entering the production chain.

ABB has developed NIR solutions that greatly accelerate the quality control of incoming raw materials which can be fully validated. Analysis is done either directly at the unloading dock or in the laboratory. NIR analysis requires no sample preparation, is non-destructive, generates no waste matter, eliminates operator bias and greatly reduces exposure to potent drugs. The analyzer can be operated by non-skilled personnel and the analysis can be performed in 5 to 10 seconds! As a result, production cycle time is reduced, highly skilled personnel are free to manage other tasks, and costly inventories are reduced.

2. NIR Solution

2.1 Theory

Near-infrared (NIR) radiation appears between Mid-Infrared (MIR) and visible (VIS) regions in the electromagnetic spectrum. It covers the frequencies between 14,286 and 4,000 cm^{-1} (700 and 2500 nm). The molecular spectra in that region are dominated by vibrational overtones and additive combinations of fundamental vibrational modes. These are much weaker and broader than the fundamental vibrations encountered in the MIR region. The most sensitive NIR bands are those derived from the O-H, C-H, and N-H stretches.

2.2 Analysis techniques

Different sampling techniques are used for solids than for liquids. Solids are normally measured by diffuse reflectance. Diffuse reflectance measures the intensity ratio of scattered light from the sample compared to the light reflected from a reference surface. Because absorptivity is lower in the NIR range, radiation typically penetrates 1 to 3 mm into the material. This gives more representative sampling than Mid-IR, which only probes the first 5 to 100 μm of a sample. The depth of penetration depends on particle size, their packing and the wavelength. The scattered radiation is detected and plotted as a function of the radiation frequency.

Liquids are usually measured in the transmission mode. This mode measures the decrease in radiation intensity as a function of the wavelength when radiation passes through the sample. Since glass is virtually transparent in the NIR region, low-cost glass containers can be used when working in the NIR range. When the NIR range is used, light can be transported over long distances (100 meters or more) in low-OH silica fiber optics. Thin polymer films and bags may also be used as sample containers in order to avoid direct contact between samples and sampling accessories. These attributes make NIR analysis easier to use than Mid-Infrared analysis.

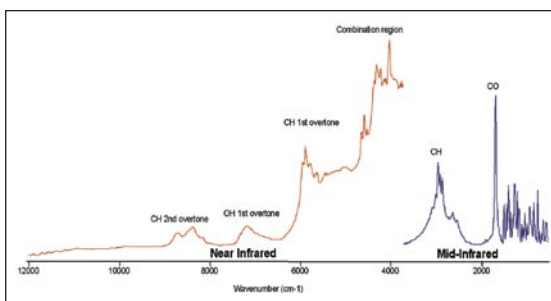


Figure 1: Mid and near infrared spectra of ibuprofen.
(The intensity scales do not reflect absolute band intensities.)

2.3 What products can be analyzed with NIR?

In order to absorb NIR radiation, a molecule must normally contain at least 1 hydrogen atom.

Consequently, most organic solids and liquids are measurable using NIR. Complex or hydrated inorganic salts can also be measured, although their NIR spectra are less specific than the spectra obtained from organic substances. Inorganic acids can be measured as a result of their effect on the water band. Packing materials and films can be analyzed in diffuse reflectance using a reflective standard located behind the film. Greasy and transparent, highly viscous materials are easily measured in transmission when melted. Emulsions and dispersions can also be measured in transmission if they contain at least 10% organic phase.

Solid salts like NaCl, KCl, KBr, etc. do not significantly absorb infrared radiation and are therefore often used as optical materials (windows, beamsplitter, etc.) in an FT-IR instrument. The spectral differences between these chemicals are best illustrated by variations in the baseline of the spectra. However, baseline variations can also be caused by differences in particle size. Therefore, identification of these salts with NIR is not advisable.

2.4 Can NIR detect contaminants?

Raw material ID is primarily used as a rapid qualitative analysis. Sample spectra are compared to the acceptance criteria set for that particular product in the database. The limit of detection (LOD) depends largely on the NIR activity and the similarity of the NIR spectrum of the contaminant with the spectrum of the desired product. In solids, NIR-active impurities between 0.5% and 5% levels can be detected. In liquids, NIR-active impurities can be detected in concentrations as low as 100-5000 PPM depending on the contaminants and acquisition time. Furthermore, water contamination in organic liquids can normally be detected in concentrations as low as 10-50 PPM.

2.5 Can NIR be used to for purity testing?

NIR is ideally suited to performing quantification of one or more NIR-active analytes in the presence of other NIR-active components. The percent determination is based on the spectral changes that occur in the spectrum of the mixture as the ratio between components changes. These changes are subsequently modeled by a chemometric algorithm. For example, a chemometric model would not work for a binary mixture with one NIR-active material and one NIR-inactive material that varied in composition. A change in spectral shape will not occur as only the NIR-active material will absorb NIR radiation.

The answer to the question, can NIR be used for purity testing, depends upon the physical state of the sample. For liquids, the answer is yes provided a sampling accessory with a defined

pathlength is used. A simple Beer's law model relating absorption to concentration can be used when the pathlength is known. Disposable glass vials, which are adequate for other quantification methods, are not appropriate for purity testing as there are small variations in pathlength between each vial.

For solids, the answer is no. Solids accessories work by either diffuse reflectance or diffuse transmission. In either case, the pathlength is never constant or known. Mathematical normalization of pathlength is thus required to provide accurate quantitative results. However, this mathematical correction will cancel the effect of varying purity on the spectral features hence eliminating the possibility of using NIR for the quantification of purity in solid raw materials.

2.6 What is the effect of temperature changes in samples?

NIR spectra are affected by differences in temperature. Although these differences are not so critical with solids, liquids require temperature control to be measured accurately. This is especially true for aqueous solutions. They are subject to substantial changes in the position of the water band if sample temperature is not controlled. In general, molecules that form hydrogen bonds, inter-molecular and intra-molecular bonds are very temperature sensitive, while C-H bands are less so. Temperature control is always required for discrimination based on viscosity.

2.7 How does the identification work?

The NIR spectrum of a material is acquired and processed through a calibration model developed for a given material using a chemometric approach. Results must meet all predefined acceptance criteria in order to be positively identified. Furthermore, the spectrum is predicted against all the calibration models of the specific materials in the database to avoid false positives or multiple identifications. Multiple algorithms are available in order to build twenty models such as: wavelength correlation, Principal Components Analysis (PCA) and, Partial Least Squares (PLS), etc. For both PLS and PCA algorithms, a set of product content-weighted eigenvectors, or factors, are generated based upon a calibration set of spectra. Each factor corresponds to a variation present in the calibration set. Linear combinations of these factors are used to reconstruct the calibration spectra. The quantity of each factor required to reconstruct a spectrum is called a score. Following reconstruction, there is a leftover called the spectral residual. Acceptance criteria can be set for the SCORE RANGE of each pertinent factor, the SPECTRAL RESIDUAL LIMIT and, for the PLS algorithm only, the PREDICTED

VALUE RANGE. In each case, the determination of these limits is based on statistical analysis of the data during model development to prove effective discrimination between different groups while maintaining acceptance of a material to its group.

ABB uses individual PLS models since the high discrimination power of this algorithm results in the lowest probability of obtaining false positive identification. A PCA based method, while perhaps the most widely implemented technique for discrimination in chemical practice, is essentially an unsupervised classification method. It makes the assumption that between-class variability soundly dominates total variability. The success of its application is dependent upon the constructed PCs accounting for a majority of between-class variability, but in practice can be challenged by within-class variability of similar chemical species among its members. In this situation, the PCA model can no longer guarantee desirable discrimination and runs the unacceptable risk of false-positive results. A second choice is soft independent modeling of class analogy (SIMCA). This approach is a preferred way in PCA discriminate analysis because it takes advantage of class memberships by building an individual PCA model for each class. The models, however, are still constructed without regard to other classes. This deficiency means that some factors that may be included in the modeling of each class, while capturing significant within-class variance, represent little discriminate information. When the sub-spaces of different classes in high-dimensional data (e.g. NIR) are very close to one another, there exists the risk of producing non-optimistic classification models due to unnecessary overlap between classes. Individual PLS models are the superior alternative for discrimination analysis because the construction of factors takes into account class membership information provided by coding of the secondary matrix. Optimization of the fitting and prediction for membership-indicating variables (i.e. wavelengths frequencies) is a way to improve the separability of the factors.

3. Operational Aspects

3.1 Instrument configuration

Two instrument configurations are available, depending on whether or not remote sampling is required. Remote sampling is used when a sample cannot be easily transferred into portable containers and brought to the analyzing station because the sample container cannot be opened due to its toxic, hygroscopic or unstable properties, to logistics at the receiving dock, or even in order to accelerate analysis.

If remote sampling is required, the ABB FTLA2000-PH30 is equipped with a fiber-optic coupled diffuse reflectance probe for solids (ACC102) and an optional temperature-controlled disposable glass vial accessory for liquids (ACC118, 131, or 132). The diffuse reflectance probe can analyze samples through polymer bags, sheeting materials or glass.



The FTLA2000-PH20 series bench-system is the preferred choice for analyzing samples that can be brought to the analyzing station. It is equipped with an in-compartment diffuse reflectance accessory (ACC101) for solids. An optional temperature-controlled vial accessory (ACC118, 131, or 132) can be used to analyze liquids in disposable glass vials, either on the auxiliary side sample compartment or in the main compartment. The advantage of this system is that it can also be used for quantitative analysis. Other accessories are available for analyzing gels, creams, pastes, and other types of samples.

3.2 Sampling options

3.2.1 Sampling of solids

Solids can be sampled using three different techniques:

1) By direct contact

This technique yields the best quality spectrum but is more difficult to implement since the accessory requires cleansing after each sample. Consequently, the possibility of cross-contamination between samples is significantly increased and the cleaning procedure must be validated.

2) Through LDPE bags or sheeting material

Measurements can be carried out through these materials in order to avoid direct contact with the sample Figure 2. These materials will absorb some NIR light (and dominate the aliphatic C-H region) but the calibration models can be made almost totally insensitive to the thickness, quality, and color (for light colors) of the polymer. This can be achieved by excluding the absorption bands of the specific plastic from the calibration. As shown in Figure 3 for LDPE, the regions $< 4500 \text{ cm}^{-1}$ between



Figure2: Sampling by direct contact

5600 and 6000 cm^{-1} (C-H 1st overtone) and between 8000 and 8600 cm^{-1} may be excluded from the calibration. Another option is to perform an automatic mathematical subtraction of a representative LDPE spectrum prior to analysis (Figure 4). Both methods can be used to distinguish most materials, but it might not distinguish like materials having minute differences in chain length or particle size.

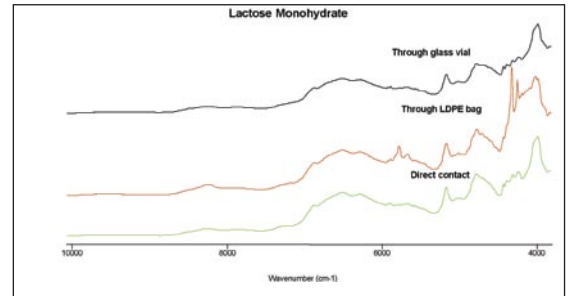


Figure 3: Sampling through a polyethylene bag with the PowderSampler (ACC101)

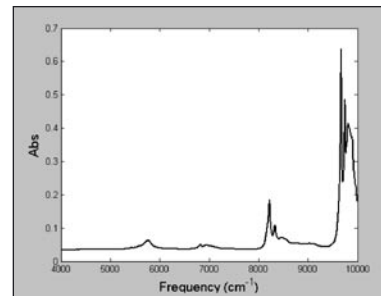


Figure 4: Polyethylene NIR Spectrum

2) Through glass vials or bottles:

Glass is virtually transparent in the NIR. The quality of spectra obtained is comparable to that obtained through direct contact. The glass should be thin (preferably 1 mm or less). Variations in glass thickness result in an offset of the spectrum that is easily compensated for by means of a flat baseline correction. The glass should also be clear or lightly colored since the NIR absorption of amber bottles is high with frequencies higher than of $6000\text{-}7000 \text{ cm}^{-1}$

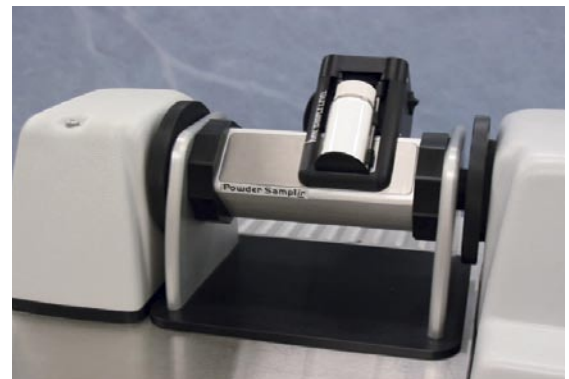


Figure 5: Sampling through glass vials with the Powder Sampler (ACC101)

3.2.2 Sampling of liquids

Disposable borosilicate glass vials of varying diameters can be used with the liquid sampling accessories (ACC118, 131, or 132). Pathlength correction algorithms can successfully compensate for variations in the internal diameter of glass vials. The repeatability of results obtained with different glass vials is comparable to the repeatability obtained when using a single vial. Disposable glass vials eliminate the danger of cross-contamination. The use of glass permits greater flexibility when sampling corrosive chemicals since glass is inert. The accessory is temperature-controlled within $\pm 1^\circ\text{C}$ to ensure the spectral reproducibility required for stringent discrimination.

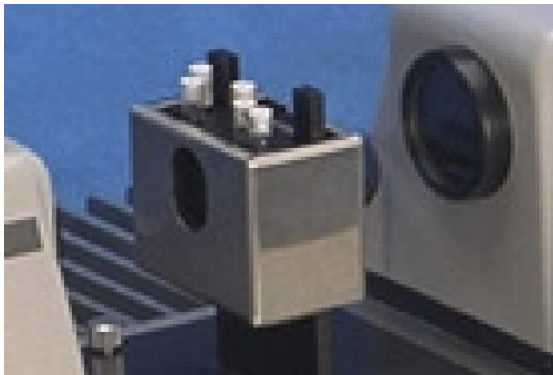


Figure 6: Sampling liquids

3.3 Instrument Validation

All ABB instruments are supplied with full IQ and OQ protocols for each of their sampling configurations. They are also completely documented for software validation and comply with the latest EP, USP, PASG and GAMP standards. Daily and monthly testing protocols are implemented in the instrument's software offering full data archiving and reporting capabilities. The complete OQ protocol can be done in 10 to 15 minutes.

3.4 Instrument Maintenance

FT-NIR instruments require minimal maintenance as compared to alternate analytical techniques. The preventive maintenance program requires replacing the infrared source once every 6 months and the laser once every 3 years. Yearly instrument revalidation is strongly recommended.

3.5 Implementing a raw material identification analyzer

3.5.1 Building Individual Identification models

3.5.1.1 Steps required to develop a calibration model

Step 1: Library Methodology (Individual versus Common Model)

The common model methodology uses a single model as the basis of the discrimination for

all materials in the library. This methodology necessitates a compromise in the selection of parameters focusing on the entire set as opposed to the members that comprise a set. The advantage of this approach is that it is computationally fast. The disadvantages are that it is less robust, requires more work to set up, and, most importantly, that adding another material requires rebuilding the entire model and performing a re-validation. To add a new material to the library database in a common model system the whole library should be revalidated to demonstrate continued specificity.

The individual model methodology uses a separate model to discriminate each material group from the others. Each model has its own unique discrimination criteria. It is the preferred method for improving discrimination power as the optimal data pre-processing/transformation and discrimination criteria are selected per group, which maximizes the inter-class separation while at the same time minimizing the intra-class variability. In addition, because each material group is calculated separately, inclusion of a new classification group or modification of a particular group does not require re-optimization and re-validation for all groups. With this approach, a model is simply built for a new group and is added to the list of existing models, or an existing model is replaced by a newly validated one, without requiring revalidation of the entire library. For most applications, the use of INDIVIDUAL MODEL is more practical than COMMON MODEL and is highly recommended by ABB.

Step 2: Spectral Similarity Analysis

To properly develop and validate an NIR ID method, two sets of samples are required: one for construction of the model and an independent one for the validation set. For calibrations using PLS, the database must consist of at least two distinct classes of material spectra in order to obtain high discrimination power between similar materials. This can be done chemically, by putting materials into groups/families based on their chemical formulas. Materials are normally classified as solids or liquids, then according to their functional groups, such as acids, alcohols, amides, amines, etc. and, finally, according to their physical properties: chain length, viscosity, particle size, etc.

Step 3: Selection of Discrimination Material(s)

For calibration of identification methods using PLS, the database must consist, at a minimum, of two distinct classes of material; the target material for identification and a discrimination material. The discrimination material serves as a basis to alter the factors to maximize the relevance of the variance that best describes the difference between the two materials. This is unlike other

multivariate algorithms (e.g. PCA) where the vectors merely represent the most common spectral variation in the data set. For example, the main source of spectral variability may originate from water content for a particular raw material. Using PCA, i.e. SIMCA, the spectral region associated with oxygen-hydrogen bonds would be considered most important in the ID model/method, and this may not be optimum for the desired level of discrimination among raw materials that also have variations in water content, which is not a distinctive property of the target material. Using PLS in combination with the appropriate choice of discrimination material would not place high loading weights on the regions associated with oxygen-hydrogen bonds unless it serves to distinguish between the two materials.

Method specificity is increased using PLS for identification because the spectral differences between two spectrally or chemically similar materials is maximized, significantly reducing the probability of a spectrally or chemically dissimilar material being classified as the target material (false-positive error).

Step 4: Selection of Samples for Calibration, Tolerance Limit Determination, and Validation

A minimum of 5 calibration target material samples and a minimum of 3 discrimination material samples are required to develop a PLS ID calibration model. This calibration set should optimally contain samples from different suppliers, different batches, etc, to ensure robustness. A second set of target material samples, covering the same variation, but which were not included within the calibration model, should be prepared. The purpose of this second set (Tolerance Determination) is to calculate the acceptable Pass/Fail limits to be instituted. The resulting calibration model and corresponding acceptance criteria are subsequently referred to as a Raw Material Discriminator (RMD). The third and final set of samples (Validation Set) should consist of both a number of target and non-target sample spectra. This test set provides verification for the specificity of the RMD as only the target samples should pass the defined acceptance criteria and all non-target samples should not. Prior to inclusion into any of the three sets, all samples should be analyzed using their respective compendial ID methods to verify their identity.

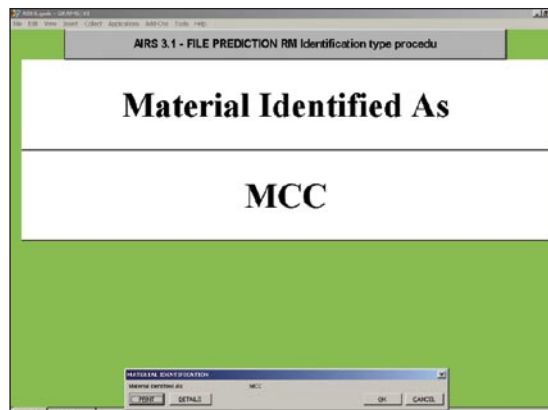
Step 5: Model Development

The first step is to select the calibration spectra of a family and to set the preprocessing options, which are almost always the same for any given identification model. A model is calculated with the chemometric package using the PLS algorithm. The quality of the discrimination can easily be seen using various plots. For each

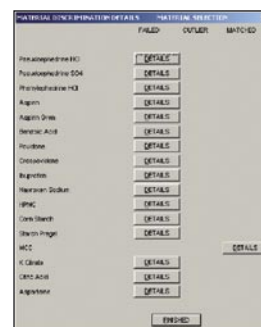
possible discrimination criterion (RANGE, SCORES and SPECTRAL RESIDUAL), a standard deviation (sigma) of the values of the tolerance limit spectra can be calculated. In each case, the acceptance criteria are normally set at a minimum of ± 3 sigma around the mean values. This ensures that, at least, 99% of the samples will fit the acceptance criteria while keeping a very low probability of false positive predictions.

Step 6: Initial Model Validation

Cross-checking a sample is carried out automatically since the spectrum is checked against all RMDs contained within the library. An example of Microcrystalline Cellulose (MCC) predicted in a library is provided in Figure 7. If a good target material validation sample fails to provide a positive result, this sample should be added to the calibration or tolerance set and recalculated. If a non-target sample is identified as a target (false-positive error), the probable cause of error may be either that the incorrect discrimination materials was chosen or the acceptance criteria were not properly set. Recalculation of the model or acceptance criteria can be performed but the models or acceptance criteria settings for all other RMDs in the library need not be changed. However, revalidation of the RMD in question is necessary. The procedure must be documented in order to comply with regulatory agency requirements.



Material Identified screen



Material Discrimination Details

Figure 7: Microcrystalline Cellulose (MCC) prediction

3.5.1.2 Estimated time required to develop models and validate 30 products

It takes approximately 3 to 4 weeks for a trained calibration developer to calibrate and validate the analyzer for 30 products depending on the complexity of the discrimination required. The steps required to build and implement the models are outlined in Figure 8.

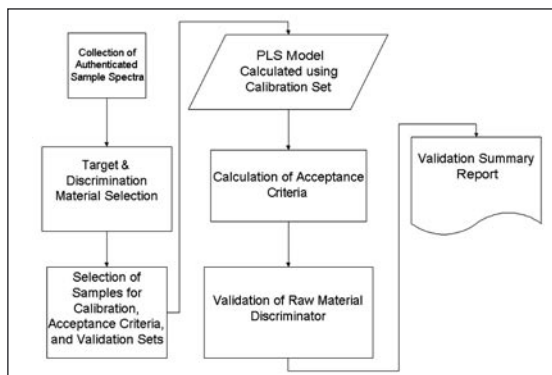


Figure 8: Steps required to build and implement a model

3.5.2 Final analyzer validation

Once the initial RMD development and validation has been performed satisfactorily, the analyzer should be put in its normal operating environment (unloading dock or laboratory). It is recommended that NIR analysis of routine incoming raw materials be performed in parallel with the standard ID analytical technique to prove its equivalence. In order to accumulate meaningful statistics, this operation normally lasts between 1 and 3 months, depending on the frequency of reception of the different raw materials. The conditions of the validation protocol should reflect the normal analyzing conditions (different operators, different suppliers, etc.). Once this is done, the analyzer is ready for implementation.

3.5.3 Model Maintenance

No model maintenance is required to compensate for instrument drift as ABB's FT-NIR instruments offer unsurpassed frequency scale repeatability (0.002 cm^{-1}), accuracy (0.04 cm^{-1}) and intensity scale reproducibility (0.002 A). Model maintenance will be required when adding a new supplier (if the initial model and discrimination settings fail to identify a good sample) or when entering a new material to the database. However, as the models are developed individually for each material, it will not be necessary to recalculate or revalidate the models for the other materials found in a given family, unless a product overlap is seen during the validation process. ABB's approach reduces the time and cost required to re-validate a method when one product is added or modified.

3.5.4 Model transferability

Development and validation of models represent the largest investment in the implementation of

an NIR analyzer for Raw Material Identification. Therefore, model transferability between instruments or after instrument maintenance is of prime importance. ABB guarantees transferability between FTLA2000 spectrometers having the same sampling accessory.

4. Performance Example

4.1 Example of discrimination of Pseudoephedrine Hydrochloride from Pseudoephedrine Sulfate

Typical raw spectra of two products in the NIR region from are presented in Figure 9. The Pseudoephedrine Hydrochloride and Pseudoephedrine Sulfate spectral differences are accentuated once the spectra have been normalized to correct for particle size related variance (Figure 10).

The predicted value plot (Figure 11) of the two-factor PLS model used for discrimination shows a strong discrimination between the two forms of pseudoephedrine. The residual plot (Figure 12) is also used for the purpose of discriminating all other chemical spectra from pseudoephedrine. A residual plot of 670 spectra representing approximately 300 different raw materials also shows a clear and distinctive separation between pseudoephedrine and all other chemical species. Other results such as scores plots can also be used to discriminate between different materials.

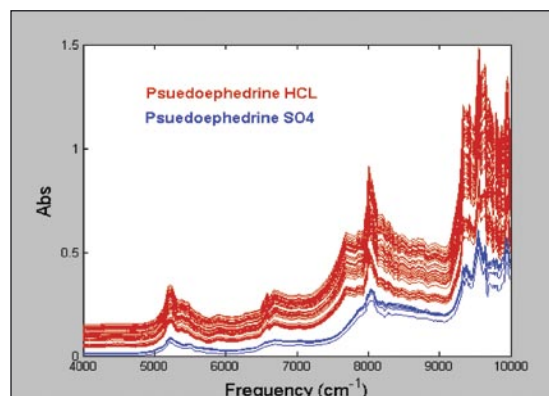


Figure 9: Raw spectra

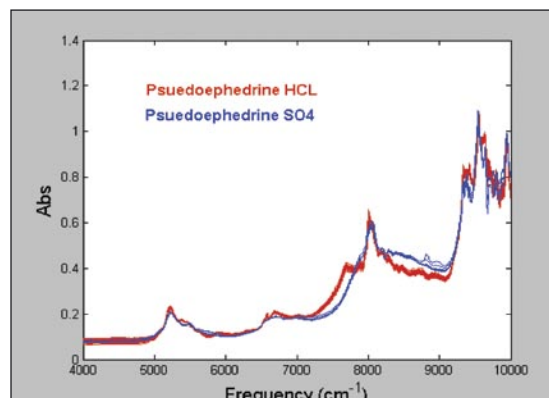


Figure 10: Normalized spectra

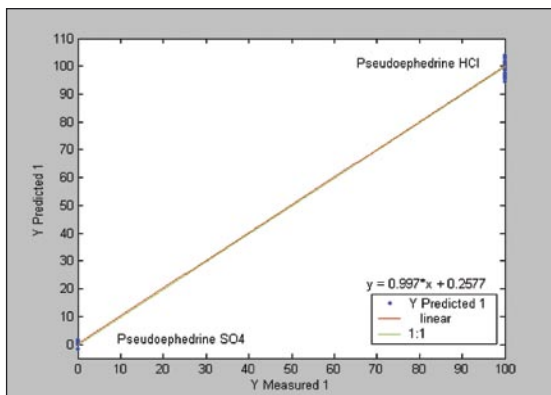


Figure 11: Predicted Value plot

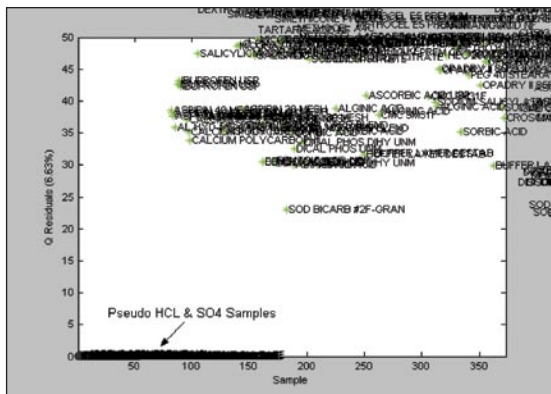


Figure 12: Residual Plot

4.2 Probability of false positive identification

Statistical calculations on the acceptance criteria based on 99.99% coverage of the population distribution obtained above yield a probability of a false-positive identification for Pseudoephedrine Hydrochloride being exponentially lower than 0.00003% (greater than 5 standard deviation separation).

The PLS Calibration of Pseudoephedrine Hydrochloride versus Pseudoephedrine Sulfate therefore ensures the lowest probability of false-positive error, that is, the error of identifying any material that is not Pseudoephedrine Hydrochloride as Pseudoephedrine Hydrochloride. The high level of specificity (low probability of false-positive), provides the justification for substitution of the NIR ID method for the full body of compendial ID test methods (see Table 1).

USP required ID tests	Alternated testing by NIR (Yes/No)
Mid-IR <197K>	Yes
Chloride Test <191>	Yes

Table 1: Pseudoephedrine Hydrochloride

5. Summary

The use of NIR analyzers for identification of raw material can significantly decrease analysis cycle time. This reduces inventory held for analysis and production cycle time. The advantages of using ABB's FT-NIR instruments are numerous. Our analyzers show high intrinsic stability that eliminates the need for model maintenance. Measurements can be carried out in a non-contact mode, which eliminates cross contamination. Models are developed using the PLS algorithm instead of PCA to obtain the lowest probability of false positive identification and the possibility to build individual models that are easier to implement, validate, and update. Model transferability is guaranteed between instruments, notwithstanding instrument maintenance.



ABB
Analytical and Advanced Solutions
 585, boulevard Charest E., suite 300
 Québec, Qc G1K 9H4
 Canada
 Phone: 418-877-2944
 Fax: 418-877-2834
 Email: ftir@ca.abb.com
 www.abb.com/analytical

Application Note B4218 2005-01

Copyright © 2004 by ABB Inc. All rights to trademarks reside with their respective owners. Specifications subject to change without notice. Pictures, schematics and other graphics contained herein are published for illustration purposes only and do not represent product configurations or functionality. User documentation accompanying the product is the exclusive source for functionality descriptions.