NIR calibration in practice

S.C.C. Wiedemann, W.G. Hansen, M. Snieder and V.A.L. Wortel

Unichema International, R&D, P.O. Box 2, 2800 AA Gouda, The Netherlands

NIR spectroscopy is growing very rapidly and chemists have to come to terms quickly with the technologies involved and their ongoing developments. In this paper, two successful NIR applications are used to illustrate the basic but important principles applied when building NIR calibrations.

Near-Infrared (NIR) spectroscopy has been growing very fast in recent years [1-2]. Both manufacturers and users have made tremendous efforts to optimize NIR techniques, in terms of spectrometer stability, as well as prediction reliability and ruggedness. However, when compared to other analytical techniques, NIR technology has been held back by its "black box" image. Simplicity of both sampling and analysis procedures, along with the development of "user-friendly" software, hides many traps from the novice.

It is well known that the NIR spectral region corresponds mainly to overtone and combination bands; Interfering peaks together with various optical effects result in a very complex spectral matrix [3-5]. Valuable chemical information must be extracted using modern mathematical tools [6]. As a consequence, developing a NIR application requires interdisciplinary skills. Despite many attempts to bridge the gap between spectroscopy and chemometrics [6-7], calibration processes remain poorly applied in the NIR field. Moreover, because NIR technology has been driven by industry needs, NIR literature has focused mostly on the description of successful and profitable implementations [8]. Novel concepts or ideas are scattered in publications of various application areas and disciplines. On the other hand, papers often convey opposing views [9], mostly based on empirical observations, making their assessment difficult for the chemist.

As a result of this lack of understanding, pitfalls are encountered from the early stages of development. NIR spectroscopy is typically a secondary technique. Thus, regression to a reference method relies on a set of samples

with known characteristics. Calibration dependence on those samples remains underestimated. In addition, proliferation of algorithms leads newcomers to believe falsely that they have circumvented the difficulties inherent to a data set (outlier or unique sample, homogeneity, nonlinearity...) by simply increasing the degree of calibration complexity. Choice for a particular calibration method is rarely mentioned in papers, and models are often built simply by "trial and error". Proper validation steps giving a realistic idea of both accuracy and precision are still neglected. These limitations are generated or amplified by time constraints governing the industrial world. Finally, calibration requirements for the future are difficult to predict [10]; the NIR methods often fail after a period of time, if not immediately. Poor understanding of the various fields involved results in an inability to evaluate the contribution of the calibration to the prediction errors and later to apply the relevant corrections.

In fact, the extreme variety of NIR applications has not encouraged a single and systematic approach. Only a few publications are concerned with such an undertaking [11]. Experience appears still to be the best way to overcome complex situations encountered with NIR spectroscopy. Complexity of calibration processes in the NIR range is the subject addressed by this paper. A few elementary concepts are thus illustrated through two different applications: The first one aims to monitor batch processes of series of ester products, while the second one concerns quality control of solid samples. Through these two particular examples, we demonstrate that the analysis target should be defined at the very early stages of the project, in terms of accuracy and precision, but also in terms of robustness. Finally, we deliberately exclude from this paper discussion on algorithms, since this topic has already been widely reported [6].

Material and methods

After cooling to room temperature and grinding, NIR analysis of **solid samples**, was carried out using a NIRSystems 6 500 spectrometer (NIRSystems Inc., Silver Spring, MD, USA), equipped with a tungsten halogen lamp, a PbS detector and a spinning sample module. Spectra were recorded at a 2 nm interval from 1100 to 2500 nm.

Liquid samples were analyzed using a ProSpec IV NIR analyzer (UOP-Guided Wave Inc., El Dorado Hills, CA, USA), linked to a Single Side Transmission probe (UOP-Guided Wave) by two ten meter single optical fibres (UOP-Guide Wave). The source of light is a tungsten halogen lamp and the detector is made of InGaAs. Scanning starts once the sample temperature has been stabilized at 65 °C (± 0.25 °C), and covers the region 1100 to 2100 nm (resolution of 1 nm).

Spectra from solid samples were used after second order derivatisation, while liquid samples were baseline corrected on one wavelength. After data treatment, spectra were exported to the Unscrambler software® (5.5/6.1, Camo AS) for investigation. Statistical calculations were executed using SAS package® (6.12, SAS Institute Inc.).

Results and discussion

The critical parameters affecting NIR models are discussed in the following paragraphs using two particular examples. Implementation of both applications has proved that NIR predictions are compatible with wet chemistry reference values. NIR spectroscopy has in fact allowed a real improvement of repeatability and reproducibility.

Importance of sample type, NIR analyzer and reference method

First priority should be given to the samples, their origin, the reaction processes involved and the reference method used. Moreover, absorption bands for the compound of interest should be investigated. In other words, NIR applications should be developed based upon a sound knowledge of chemistry and spectroscopy.

Let's start with sample conditions. By definition, calibration samples should be representative of the future samples to be analyzed. However, they are subject to many controlled or uncontrolled factors. For instance, the two applications addressed here use raw materials of natural origin and carbon distribution of the fatty acid involved may change with the supplier, the origin or the season. Conversely, moderate variations in process (temperature, pressure...) are not an issue here, since all measurements are carried out "at-line". NIR spectroscopy allows minimization of sample preparation. However, although the operator workload can then be drastically reduced, the samples to be analyzed often present a complex matrix, leaving to the development chemist the task to list major and minor components, as well as the possible contaminants. Although NIR spectra cannot be fully characterized, band origin should be studied and possible interfering constituents should be recorded.

Returning to the example of the esterification process, NIR models were built to follow both acid and hydroxyl values decrease; Doing so, calibrations obviously also focus on acid/alcohol ratio in the reactor. Sample collection was thus carried out during different batches, to avoid artificially including an extra correlation between these two major reagents. In addition, water content was systematically checked; In fact, high levels of water affect OH bonding [12], and therefore both position and intensity of the bands used by the algorithms (Fig. 1): the first OH overtone region (1 400 – 1 500 nm) and the OH combination range (1900 – 2 100 nm). Eventually, additional outlier detection can be performed by simply checking the water absorption band around 1925 nm.

Besides chemical interferences, possible causes of physical interference should be investigated. Thus, when dealing with our liquid esters, small temperature variations of about 2 °C may lead to prediction errors for both acid and hydroxyl values up to 1% [13], due to a combination of density changes and effects of hydrogen bonding. Rather than increasing the calibration complexity by including the temperature as a variable, a sampling – measuring unit has been developed to enable automatic control of the sample temperature prior to scanning with a precision of 0.5 °C [13]. Similarly, in order to minimise the effect of particle size, reflectance NIR measurements of our solid samples required optimization of the preliminary grinding procedure. As
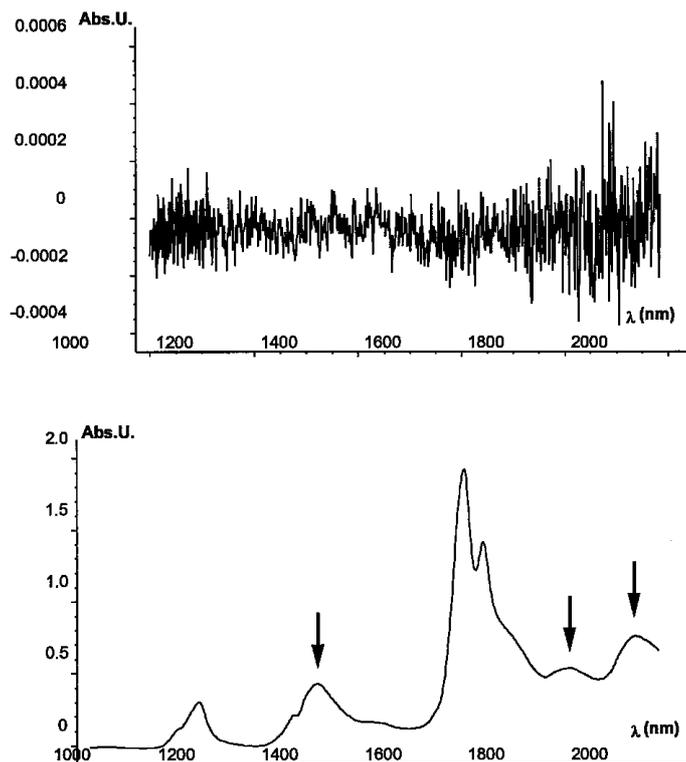
Figure 1. Instrumental noise (air spectrum) and selection of absorption bands in the ester spectrum for hydroxyl and acid number predictions.

expected, irregularity of sample surface and inhomogeneity of solid structure resulted in large light scattering effects, which could not be simply corrected by introduction of a spinning cell or the usual mathematical treatments, such as derivatisation [14-15].

Tremendous improvements have been made in reduction of instrument noise as well as wavelength accuracy. As a consequence, instrument error can be neglected when compared to the reference method error and "regression theory" – assuming no error in the X variables [7] – can be applied with confidence. However, simple inspection of our "noise spectrum" (air sample, Fig. 1) shows differences in instrumental noise as a function of the wavelengths, resulting mainly from the variation of the detector sensitivity. The high noise level in the combination range led us to avoid, where possible, the wavelengths beyond 1900 nm; Despite the loss in resolution between bands related to acid, water and alcohol, the models have thus been built on the OH first overtone only. The resulting gain in robustness is especially valuable when transferring calibrations to other instruments.

Finally, the reference method may in some practical situations deviate from the spectral measurements; hence, the calibration limits should be stated at the beginning. Thus, for OH number determination, one common problem encountered with the titration method is the lack of solubility of the analyte leading to erroneous reference data. On the other hand, publications relate discrepancies of wet chemistry methods with some specific substrates [16-17]. In both cases however, absorption of the relevant bands will still follow

Beer's law (as long as it remains within the detector linearity range).

A good data set

When building a calibration set, the three following parameters should be taken into consideration: 1. the number of samples; 2. the concentration range to be covered by the samples; 3. the distribution within this range. Common sense will usually dictate to collect as many samples as possible: a large number should in fact ensure that the model covers a large variation range and the samples are uniformly distributed within this range (in term of all possible constituents or other parameters) [6,7,18]. Hence, models for monitoring of esterification have been built by systematically collecting samples every hour from three successive batches. When possible, validation samples have been collected from a later campaign. The operation has been repeated for each ester code, being completed over a period of six months, due to the production schedule.

However, both cost and time spent in obtaining a high number of calibration samples is often not affordable. Therefore, many different procedures have been reported by the literature for building a calibration set of optimum size. When dealing with samples of natural origin, usual designs (such as mixture designs) remain difficult to apply, since the exact composition of these samples cannot be reproduced. A common strategy is therefore to select samples on the basis only of the spectral responses [18-21].

Moreover, a large data pool does not systematically lead to successful models, as illustrated by our "solid application". For example, several thousands of samples were systematically scanned after determination of water and free fatty acid concentration by the existing reference method and were therefore available for calibration. However, as shown by the fatty acid distribution in the resulting histogram (Fig. 2), this selection favours very much the extreme ranges to be calibrated and does not lead to the desired central cluster. Because a large error in the reference method has occurred, a reasonable regression becomes difficult to make. Collection of new samples to "fill-up" the gap would require too much time, and splitting into two
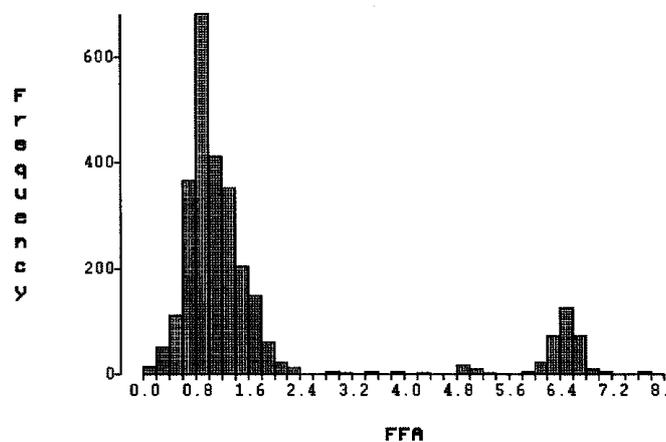

Figure 2. Using a histogram plot to evaluate the distribution of free fatty acid concentration after systematic collection of 3000 samples.

classes (as explained in the following paragraph) would restrict the calibration range, with the risk of amplifying the error of the reference method. Therefore, some samples at the extremes were deleted in order to achieve a better spread of the spectra. The selection was carried out by using SAS statistical tool®, ensuring sample diversity in terms of raw material origin, water percentage, level of other components and sampling time.

Finally, all collected samples and their corresponding spectra should be carefully inspected [7] to detect possible outliers. Even if they cannot be used to build the model, they should be kept with the data set, since they can help to evaluate later the "outlier detection".

Data pre-treatment

Multivariate techniques enable exploitation of complex spectra by enhancing both selectivity and reliability. They are suitable for handling additive interferences and practice has shown that they can even overcome some types of curvatures. However, multiplicative effects such as light scattering or path length variations requires specific pre-processing. Moreover, preliminary linearization treatments often facilitate modelling work, by simply compelling the user to face complex situations step by step.

Wavelength selection should be considered as part of the matrix construction and thus should be included into the experimental design leading to the calibration set. However, it is often carried out simultaneously with the data pre-treatment, because of its obvious interactions with the other data processing techniques [10,22,23]. Although literature on wavelength selection is considerable [24-27], resulting predictive ability is rarely evaluated in terms of robustness [10,28].

Data pre-treatment should be chosen carefully in order to prevent any increase in complexity. Smoothing can be applied when drawing information from the neighbouring wavelengths helps to stabilize the model. However, an experimental design on our ester data showed that present and even future predictive ability of the model could not be significantly improved by smoothing; in fact, the original spectra were obtained with already a very good signal to noise ratio. Furthermore, the regression method should be again taken in consideration. Thus, smoothing brings often more advantages in association with MLR method than with PLS regression.

None of the known mathematical data treatments can provide a general solution. In general, simple additive effects (baseline shift) can be removed by "baseline correction". Multiplicative differences resulting from particle size, packing variations or path length differences can often be corrected by "derivative" or "Multiplicative Scatter Correction" [28,29]. Note that these data processing have been allowed by the tremendous improvement of the instrumentation noise. However, when the study aims to investigate parameters such as the effects of sample thickness or particle size, original raw spectra should be used without any preliminary transformation [30-31].

In the case of non-linear modelling, splitting into classes remains the best way to guarantee a good linear fit between spectral and reference data [32-34]. However, the number of calibrations may increase rapidly, requiring maintenance of

Figure 3. Two first-factors scores resulting from PCA, underlying the different ranges in hydroxyl number of two esters from the same "family".

many additional models. Let's take again our ester application; Spectral and wet chemistry data (hydroxyl number, water content) characterizing two esters of the same "family" have been collected simultaneously. Figure 3 shows the score plot resulting from Principal Component Analysis (PCA); most of the spectral variance (99%) can be depicted by the two first Principal Components (PCs) and corresponds to variation in hydroxyl values (PC1) and water content (PC2). This plot shows two distinct clusters, corresponding to the two ester types. Therefore, although these two products are made from similar raw materials, they cover ranges too distinct to be calibrated together. On the contrary, a single model is presently used for hydroxyl prediction of a group of eight esters from another production line. In fact, the PCA shows five distinct groups distributed along PC2, which depicts variance of CH, CH2 absorptions (Fig. 4). However, selection of wavelengths related to OH bonds (first OH overtone between 1400 and 1550 nm and

Figure 4. Score plot resulting from PCA on a group of 8 esters, which differ by their carbon chain; The second PC accounts for CH, CH2 variance.

combination range higher than 1850 nm) eliminates any structure related to ester raw material and thus allows us to build a single model for prediction of hydroxyl number for all products. In addition to facilitating calibration maintenance, this strategy enables analysis of contaminated samples with slightly different fatty acid distribution.

Number of principal components

Although comparison of algorithms is not addressed here, overfitting problems may be faced with all methods. The question "How many principal component (PCs) should be used?" arises especially when realistic validation samples are not available. An independent test set should consist of samples from a later batch, analyzed if possible some time after calibration.

When deciding the number of PCs to be used, the analysis target as well as the chemistry involved should be kept in mind. Even when the sample matrix is complex, only a few PCs should generally be used if the variable to be modelled is related to a major constituent: Hence, prediction of hydroxyl value in our esters is a good example, since this variable undergoes large variations during the reaction, resulting in significant changes in the corresponding bands (1400 – 1550 and 1900 – 2100 nm, Fig. 5). A relevant model can then be built with two PCs, while the Root Mean Square of Prediction (based on a test set validation) remains in the order of the reference method precision. However, choice of variable to be modelled must be judicious. Let's take the example of a particular ester, for which the balance between alcohol and acid had to be calculated; Direct cali-

bration of this ratio between alcohol and acid number required two more PCs than the parallel model for hydroxyl and acid number separately. In fact, absorptions in the ester spectra are related only to those two variables and not directly to their ratio.

On the contrary, many PCs may be needed if the constituent under interest is present at low concentrations. Variations in intensity of the related absorption bands remain insignificant, so small variances cannot be extracted by the few first PCs despite the use of a PLS algorithm. Thus, prediction of fatty acid percent for our "solid samples" requires up to seven PCs (Fig. 5). In such a situation, relevant information is masked mainly by large variations of the water bands and must be discriminated carefully from the noise. This is where new visualisation tools available in chemometric software prove very valuable.

Conclusion

To summarise, a successful NIR application requires the chemist to understand fully both the complex sample matrix and the reference method. Of course, the potential advantages and disadvantages of NIR instrumentation should also be investigated thoroughly and sampling devices optimised. A comparative study of algorithms will only be profitable once these problems have been solved. Because of the diversity of NIR applications, simple "recipes" cannot be applied, and a good general understanding of the problem remains the best weapon to face practical difficulties.

In our opinion, "ready-made calibrations" should be avoided. In fact, a proper maintenance does not only rely on regular instrument checking, but also requires procedures ensuring model validity. Calibration updates such as "moving models" can only be carried out efficiently, if the analyst has been involved early enough to master the technique.
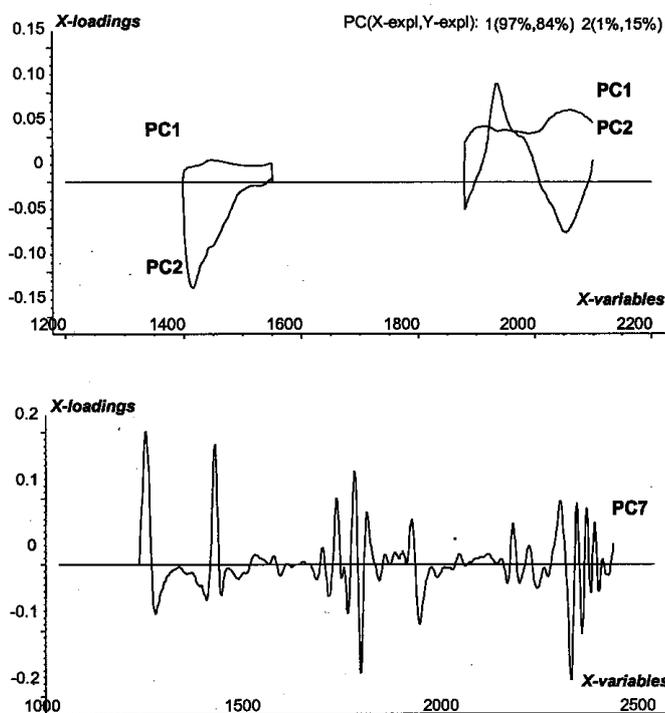
Acknowledgements

Figure 5. Using loading plot to distinguish information from noise; The calibration of hydroxyl number related to the major constituents alcohol and acid requires only 2 PCs, while the prediction of free fatty acid percent in our solid samples is based on 7 PCs.

References

1. Mc Clure, W. F. *Anal. Chem.* **1994**, *66*, 43-53.
2. Noble, D. *Anal. Chem.* **1995**, 735-740.
3. Weyer, L. G. *Appl. Spectrosc. Rev.* **1985**, *21*, 1-43.
4. Kradjel, C. *Fresenius J. Anal. Chem.* **1991**, *339*, 65-67.
5. Downey, G. *Comput. Appl. Mol. Spectrosc.* **1995**, 217-235.
6. Martens, H.; Naes, T. in: Multivariate Calibrations, John Wiley & Sons, Chichester, 1989.
7. Mark, H. in: Principles and Practice of Spectroscopic Calibration, John Wiley & Sons, New York, 1989.
8. Stark, E. in: NIR Spectroscopy: The Future Waves, NIR Publications, Montreal, 1995; pp 700-713.
9. Rosenthal, B.; Williams, P. in: NIR Spectroscopy: The Future Waves, NIR Publications, Montreal, 1995; pp 1-5.
10. Wortel, V. A. L.; Hansen, W. G. in: NIR Spectroscopy: The Future Waves, NIR Publications, Montreal, 1995; pp 306-315.

11. Mobley, P. R.; Kowalski, B. in: NIR Spectroscopy: The Future Waves, NIR Publications, Montreal, 1995; pp 201-208.

12. Reeves, J. B. in: NIR Spectroscopy: Making Light work, Weinheim, V. C. H., 1992; pp 99-104.

13. Wiedemann, S.; Wortel, V. A. L.; Hansen, W. G., submitted for publication in the Proceedings of the 8th International Conference on Near Infrared Spectroscopy, Essen, 1997.

14. Bull, C. R. *Analyst* **1991**, *116*, 781-786.

15. Olinger, J. M.; Griffiths, P. R. *Appl. Spectrosc*. **1993**, *47*, 687-701.

16. Hartman, L.; Lago, R. C. A.; Azeredo, L. C.; Azeredo, M. A. A. *Analyst* **1987**, *112*, 145-147.

17. D'Sa, B. A.; Verkade, J. G. *J. Org. Chem.* **1996**, *61*, 2963-2966.

18. Fearn, T. in: Near Infrared Spectroscopy - Bridging the Gap between Data Analysis and NIR Applications, Ellis Horwood Limited, New York, 1992; pp 61-66.

19. Davies, T. *Spectrosc. Eur.* **1996**, *8*(4), 27-29.

20. Isaksson, T.; Naes, T. *Appl. Spectrosc*. **1990**, *44*, 1152-1158.

21. Ferre, J.; Rius, F. X. *Anal. Chem.* **1996**, *68*, 1565-1571.

22. Jouan-Rimbaud, D.; Massart, D. L.; de Noord, O. E. *Chem. intell. Lab. Syst.* **1996**, *35*, 213-220.

23. Wu, W.; Walczak, B.; Massart, D. L.; Prebble, K. A.; Last, I. R. *Anal. Chim. Acta* **1995**, *315*, 243-255.

24. Brown, P. J. *J. Chem.* **1992**, *6*, 151-161.

25. Centner, V.; Massart, D. L. *Anal. Chem.* **1996**, *68*, 3851-3858.

26. Brenchley, J. M.; Horchner, U.; Kalivas, J. H. *Appl. Spectrosc.* **1997**, *51*, 689-699.

27. Bangalore, A. S.; Shaffer, R. E.; Small, G. W. *Anal. Chem.* **1996**, *68*, 4200-4212.

28. De Noord, O. E. *Chem. Intell. Lab. Syst.* **1994**, *23*, 65-70.

29. Honings, D. E. in: NIR Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications, Ellis Horwood Limited, New York, 1992; pp 109-118.

30. Haanstra, W. G.; Hansen, W. G. et al., submitted for publication in Applied Spectroscopy, June 1997.

31. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Gonzalez, F.; Pous, R., in: NIR Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications, Ellis Horwood Limited, New York, 1992; pp 401-406.

32. Naes, T.; Isaksson, T. *J. Chem.* **1991**, *5*, 49-65.

33. Naes, T. *J. Chem.* **1991**, *5*, 487-501.

34. Hansen, W. G. in: NIR Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications, Ellis Horwood Limited, New York, 1992; pp 29-34.