Original Article

JNIRS

Journal of Near Infrared Spectroscopy
25(2) 85–90

DOI: 10.1177/0967033517702395
journals.sagepub.com/home/jns

SAGE

Tutorial: Items to be included in a report on a near infrared spectroscopy project

Phil Williams[1], Pierre Dardenne[2] and Peter Flinn[3]

Abstract

There are nearly 40 items that should ideally be reported when an NIR (near infrared) spectroscopy project is completed, either as a report or as a scientific paper. However, in our reading of the extensive literature, many of the papers presented or published report no more than 6–10 of these. The purpose of this tutorial is to indicate all of the items and the reasons for reporting them. Most of the items that need to be reported are important for anyone who seeks to duplicate the type of application and methods reported in a peer-reviewed journal article for their own work. Practically, all of the items are significant to any worker if the eventual objective of their work is to extend it to the level of industrial application. The tutorial will summarize these items, and give some explanation for their inclusion. The tutorial should be useful to potential authors, as well as to reviewers.

Keywords

Calibration, essential items, NIR spectroscopy, reporting, validation

Received 9 March 2017; accepted 9 March 2017

Introduction

There are two main types of calibration. One is a feasibility study, in which a relatively small number of samples are assembled, scanned, submitted for reference data, calibrations developed, and the results published as a scientific paper, or presented at a conference. That type of calibration is the basis of most of the thousands of papers that have appeared in journals and conference proceedings. Usually, no attempts are made to determine whether such calibrations actually work, and there is no real need, because the objective, that of demonstrating and proving the feasibility, has been achieved.

The other type is the calibration that is developed for use in industry. There are two sub-divisions of this type of calibration. The first is intended for application to every sample that arrives at the laboratory for analysis, covering the entire expected range of composition and other variables. An example would be the testing of deliveries of grain for protein or moisture contents. Such calibrations are typically in daily use in industry. The true test of the effectiveness of such calibrations surfaces when a set of early samples, having been tested immediately after application in day-to-day analysis, are subjected to the ubiquitous statistics, SEP, r^2 and the RPD.

The second sub-division is the calibration that is developed for quality control, for example in flour mills or feed manufacture, where consistency in the product is essential. For this type of application, minimum variance is expected in the samples subsequently tested by the calibration. For practical purposes, the quality control staff are really only interested in the predicted results themselves, in order to ensure that specifications are being met, e.g. 12.2% protein $\pm$0.1% in the flour. Of the usual statistics, the SEP is of certain value, but the r^2 and RPD are not applicable because of the absence of sufficient variance. But this type of calibration is equally important. For example, a big flour mill can produce 1000 tons of flour an hour and it is imperative that the flour is meeting bakers' specifications all along. Both of these sub-divisions call for extensive work, including assembly of appropriate samples, and the steps for development of such calibrations lies outside the scope of this tutorial. Progress in NIR spectroscopy application, as is the case in other scientific endeavours, relies on keeping detailed records at every stage of the work. The items

[1]PDK Projects Inc., Nanaimo, British Columbia, Canada
[2]CRAW, Gembloux, Belgium
[3]NIR and Feed Evaluation, Victoria, Australia

Corresponding author:
Phil Williams, PDK Projects, Inc., 5072 Vista View Crescent, Nanaimo, British Columbia V9V 1L6, Canada.
Email: philwilliams@pdkgrain.com

that we regard as being essential to progress are listed and briefly described in appropriate groups.

Group I. Introduction, sample description and reference methods

1. For what purpose is the calibration being developed?
2. Material, and form of material, e.g. solid, liquid, slurry
3. Source of the samples
4. Number of samples collected
5. Sampling method, including sample size
6. Method of sample storage
7. Method of sample preparation
8. Reference method(s) with units
9. Moisture basis for reporting: e.g. dry matter (DM), as-is, or some other moisture basis
10. Standard error of the reference method(s), i.e. the standard error of the laboratory (SEL), or the standard error of the test (SET)

Item 1. The purpose. The need for the application should be the first item to examine, when consideration is being given to undertake an NIR spectroscopy application. It should indicate why the work was undertaken and should include the value or potential value, to industry. It should indicate the reasons underlying the decision to use NIR spectroscopy, e.g. the need for instant results, improvements in the costs of operation, etc. These first points should be covered in the Introduction to a report or article, and should also explain the value of using NIR spectroscopy in place of conventional methods.

Item 2. The form of material. This should include whether the material is fresh or mature, dried, whole-grain, fibrous, etc. The physical nature of the samples should be described, including whether they are fresh, ground, chopped, liquid, slurry, or otherwise. The nature of the material is usually revealed in the title of the article, e.g. grains, forages, manure, etc. Alternatively, the title should explain the purpose of the report, such as ''A new algorithm for prediction of film thickness, using NIR spectroscopy''. A spectrum of a typical sample can be included here.

Items 3–7. The source and number of samples and sampling method. These items concern the samples. The integrity of the samples used in the work is critical. This involves identification by the operator of all of the sources of variance likely to influence the spectra. Next in importance is to identify the source, or sources of samples that cover the sources of variance, and also the extent to which samples from these sources can consistently be supplied. Temperature, water content, and particle size and particle characteristics are important and recurring variables, but other sources of variance are specific to certain applications. The assembly of the samples must include and replicate all of the sources of potential variance, and the total number of samples assembled should be stated. Replication means that separate samples that carry the sources of reference are taken at different times, ideally from different sources, and not that the same sample is simply scanned more than once, or that more than a single sample is withdrawn from the same source at the same time. Item 5 is the sampling technique which comprises the method of sample withdrawal and sample size. The method of withdrawing the samples should be described, especially if samples are to be withdrawn from a process stream in conjunction with on-in-line analysis.

The method of storage of the samples can affect the composition, particularly the moisture content, and thus impact the spectra. Samples can be stored at 5°C for short periods (up to three to four days). It is not advisable to freeze grains and other cellular materials because the expansion that accompanies changes from water to ice may affect the spectra. This is particularly important for fresh samples, which should be scanned at the time of withdrawal, or as soon as possible thereafter. Such fresh samples that have been scanned can be frozen after scanning to protect them from change in composition before reference analysis, provided that this does not impact the spectral response. The containers in which samples are stored should be reported. The samples should be stored in containers or plastic bags that protect the material from alteration in composition, by moisture change, oxidation, or other agency.

Item 7. Sample preparation. One of the main attributes of NIR spectroscopy is the need for minimal and sometimes no sample preparation. Removal of obvious items that are not part of the material to be analyzed, such as foreign material from grain, is the first step in the analysis. Where necessary, the sample preparation for NIR spectroscopy analysis usually involves size reduction by grinding or chopping, both of which affect the spectra. The type of grinder and method of grinding, including particle size, should be reported. Sample preparation steps also include blending of materials such as flour streams in flour mills, and animal feed mixes, which contain ingredients that differ widely in physical make-up. Agitation of liquid materials, such as manures or sewage sludges, which contain suspended particles, should be done immediately before scanning or weighing of the sample before the reference analysis.

Items 8–10. The reference analysis. The methods used should be given, together with the associations that have certified the methods (such as the AOAC), the reference number of the method, and the units by which reference analysis is documented and reported, including the moisture basis, e.g. ''as received,'' dry matter basis, or specified moisture basis. Arbitrary or

non-certified methods should be described in detail. Item 10: the standard error(s) of the reference method(s) used are essential pieces of information. These are sometimes referred to as the SEL, or the SET, the standard errors of the tests.

Group II. Instrument and data acquisition

11. Instrument(s) used, including wavelength range and increments, and degree of resolution for Fourier transform-near infrared (FT-NIR) instruments
12. Number of replicate scans (repacks)
13. Method of sample presentation to the instrument, e.g. type and size of cups, rotating cell with dimensions, open hopper, etc.

Item 11. The instrument used, including its specifications are necessary information. The make (manufacturing company), type (e.g. scanning spectrometer, diode array, etc.) and model, of the instrument should be given, together with the wavelength range. The increments at which spectral data are recorded during scanning vary among instruments from as low as 0.5 nm to over 5 nm. Some instrument software, where the scanning interval is small, includes a system for ''binning'' of incremental spectral data before smoothing and derivative development. The degree of resolution should be reported for FT-NIR instruments. The number of sub-scans averaged per spectra should be recorded. The higher the number of sub-scans averaged affects the integrity of the final spectrum and the time per test. Usually, 25 sub-scans are adequate.

Items 12–13. The number of re-packs (as distinct from re-scans) should be stated. The method of sample presentation to the instrument is essential. Details should include the size and design of sample-presentation cells. The method of sample presentation, other than by the familiar sample cell, should be described, for example if the sample is to be scanned from above or below, the type and size of sample cell, whether the cell rotates, whether interactance is used and its dimensions, or whether the analysis is to be carried out continuously online. In the case of transmittance scanning, the path-length should be stated. Sample presentation for online applications to liquids should indicate whether the analysis is to be carried out in reflectance, transflectance, or transmittance mode. The temperature range during the period over which samples for calibration are withdrawn should be recorded, because of the importance of temperature in applications of NIR spectroscopy. Samples for calibration development should be taken throughout the normal time during which the instrument will be used for day-to-day analysis. This will incorporate fluctuations in ambient temperature, dust, relative humidity, and other factors that could add sources of variance to the spectra. Whether the instrument is to be used for indoor or out-of-door analysis should be stated. Fluctuations in operating conditions will be more extreme for out-of-door applications. Where NIR spectroscopy is used for analysis of samples of slurries and other semi-solid materials, the method of cleaning the sample cells between samples should be described.

Group III. Calibration

14. Mean of the reference data
15. Median of the reference data
16. Standard deviation (SD) of the reference data
17. Method of developing calibration model, e.g. multiple linear regression (MLR), partial least squares (PLS), support vector machines (SVM), artificial neural networks (ANN), or other method
18. Method of calibration evaluation, e.g. cross-validation (CV) or test-set evaluation
19. The (selection) method used if a test-set is selected from the original population of samples
20. Number of samples used for calibration and testing (evaluation) if step if the test-step procedure 19 is followed
21. The method used if CV is undertaken (e.g. ''leave one-out'', size of groups to be left out, cycling, blocks, random)
22. Mathematical data pre-treatment, including scatter correction system and derivative types (gap-segment or Savitsky-Golay and their parameters)
23. Number of outliers removed, including the reasons (e.g. spectral or other outliers)
24. Wavelength ranges used in calibration model
25. Number of MLR terms, PLS factors, SVM parameters or ANN architecture (layers and nodes)
26. Standard error of calibration and standard error of cross-validation (SECV or RMSECV), or root mean square standard error of prediction (RMSEP) if a test-set is split off

Items 14–16 concern the statistics attained as a result of application of the reference method(s) to the entire population, as distinct from item 10 which presents the standard error of the reference method in terms of the SD of the results of application of the method to a single sample. The results of reference testing are the criteria of comparison by which NIR spectroscopy testing is appraised. The mean, median, and the standard deviation of the reference data are also useful information concerning the composition of the samples.

But why the median? The median provides extra information about the distribution of reference data within the sample population. In a Gaussian distribution of data, the mean and the median should be similar. A sizeable difference between the mean and median indicates non-Gaussian distribution of the data for the constituent or parameter. If the SD is very high, relative

to the mean this may indicate that a small proportion of the population is higher or lower than the rest of the population. For example, if a population of wheat has a mean of 14.05% protein and a SD of 2.6, but a median of 14.57%, this indicates that a small proportion (e.g. up to 15%) of the population is much lower than the rest. Similarly, if the population of wheat has a mean of 12.44% protein and a SD of 2.6, but a median of 11.80% the implication is that a small proportion of the population is much higher than the rest.

Item 17. The method of developing the calibration model, e.g. MLR, PLS regression (PLS), SVM, ANN, or other method should be stated. Where proprietary software is used, the version should be stated.

Items 18–21. The methods for evaluation of the calibration usually involve CV or the test-set method. With a population of samples much larger than 150, the test-set system is often preferred to CV for the evaluation of the calibration model. The method of selection of the test-set method should be recorded (item 19). It includes the step of setting up both calibration and validation sample sets, often from the original population. For test-set evaluation, a useful system for selecting the test-set is to sort all of the samples of the population for the constituent or parameter, then select every fifth sample for the test-set. This guarantees even distribution of reference data in both sample sets, regardless of the source(s) of the samples. The system should be repeated for all constituents and parameters, and the numbers of samples in the calibration and validation sets should be given. Most NIR spectroscopy workers use variations of this method for test-set evaluation. The numbers of samples used in, respectively, the calibration and validation sample sets should be stated. The most authentic assessment of a calibration comes when the calibration is first applied to actual day-to-day analysis.

CV is recommended for the evaluation of any calibration models based on small sample sets, e.g. up to 100–150 samples. The methods of CV include one-out or full CV, wherein samples are omitted one-by-one, or group CV, in which groups of samples are omitted. This is key information, and should be stated. Full (one-out) CV is recommended for small populations of up to 150 samples. It becomes cumbersome and time-consuming for large populations of 500 or more and does not add to the value of the CV. Group CV by random selection, even small groups, adds the variance within the group to the overall variance and complements the integrity of the CV. The method of group selection should appear, e.g. number of samples per group, random selection of groups, etc. The best model from CV again has the highest r^2, lowest standard error of CV (SECV), and highest RPD. If CV is to be used, there is no need to sort or subdivide the samples, because all of the samples will be used.

Item 22. Mathematical data pre-treatment includes development of derivatives. The order (e.g. first, second, or higher) should be stated, together with the dimensions of the derivative (degree of smoothing, and segment size). The ''segment'' or ''gap'' is the number of wavelength points included in the segment. For example, the WinISI software defines a derivative by numbers such as 144, where 1 indicates the derivative (in this case first derivative), the first 4 gives the number of wavelength points between the initial and final wavelength point of the derivative (the ''gap''), and the second 4 gives the number of wavelength points over which the spectra have been smoothed (the ''segment'').

An alternative form of expressing derivative (mathematical) pre-treatments is to use D to describe the derivative, G to describe the Gap (the distance between points used to develop the difference), and S_1,S_2, etc. to indicate the number of wavelength points used to smooth the data (the ''segment''). Operators usually optimize the derivative dimensions, but it is only necessary to record the final dimensions (i.e. the dimensions that resulted in the most acceptable statistics) in a report or an article. Where scatter correction is applied the form should be stated, e.g. straightforward multiplicative scatter correction (MSC), standard normal variate, etc. If samples are to be withdrawn over a period of several days, sorting by date can provide a practicable grouping for CV.

Item 23. Outlier removal. The number of outliers removed is crucial. The main purposes of NIR spectroscopy analysis are to obtain consistently accurate results for quality control, or to develop a reliable method for use in daily operations, not to simply arrive at an attractive set of statistics for publication. Ideally, there should be no outliers. In the use of NIR spectroscopy for day-to-day analysis, there are no outliers, because everything has to be analyzed. The number of outliers is usually quite small. Removal of up to 2% of a population is acceptable, particularly in large populations of samples (>250). The reasons for such removals should be included. If plotting reported against predicted results shows that there are a lot of outliers in a calibration (10% or more), it is unlikely that all of these are outliers. A likely source of the excessive variance is in the reference data – again, the importance of knowing the error of the reference testing cannot be overemphasized.

If a large number of outliers appear during a new application, they will likely continue to be a factor in future application. If one is working on a calibration for a completely new commodity-constituent combination, and the calibration results are poor, but there is confidence in the reference data results, NIR spectroscopy may not be applicable to that area of work if accurate analysis is required. It may still be useful if the NIR spectroscopy technique is intended as a

screening method to reduce the volume of the more expensive chemical or physicochemical analysis.

Items 24–25. The number and location of wavelengths used in MLR calibrations are useful information. For the most part, three to four wavelengths usually enable a reliable calibration. If more than six wavelengths appear to be necessary, the exercise should be repeated with new sample sets. If this results in the same number, and with approximately the same wavelengths, the calibration is probably reliable. The wavelengths used in MLR or PLS calibrations, together with their assignments can be reported at the discretion of the author, with reference to an accepted text or texts, e.g. Workman and Weyer.[1]

The number of PLS factors is a measure of the potential effectiveness of the calibration. In practice, most successful PLS calibration equations indicate no more than six to eight factors. If the system uses 12–15 factors, this usually indicates some uncertainty in the computation. On the other hand, if the system uses only 1 PLS factor, the exercise should be regarded with caution and should be investigated. In view of the concepts of principal component regression (PCR) and PLS, it is unlikely that all of the variance in spectral and reference data is explicable by a single factor.

Item 26. The standard errors of calibration and CV, SEC (RMSEC) and SECV (RMSECV) should be given. The SEC that is finally accepted after optimization of wavelength range, mathematical data pre-treatment, and scatter correction should be similar to the SEP (RMSEP).

Group IV. Independent validation (samples not included in the calibration process)

27. Standard deviation of the reference data for the validation samples (SD_{VAL})
28. Root mean square standard error of prediction (RMSEP)
29. Standard deviation of the prediction error (SEP corrected for bias)
30. Coefficient of determination in validation (r^2_{VAL})
31. Regression coefficient, or slope (b_{yx} where Y are the reference and X the predicted values)
32. Intercept (a)
33. Bias
34. Ratio of SD to SE (RPD) in validation or CV (SDcal ÷ SEP)
35. Average of the global Mahalanobis distances (GH) where possible
36. Average of the neighbourhood Mahalanobis distances (NH) where possible
37. Standard error (SET) of the final NIR method (reproducibility)

Item 27. The SD of the reference data for the validation sample set should be stated if the validation sample set is a completely new set of samples that have not been previously associated with the calibration development. It should be essentially identical to item 16, which should be used for calculation of the RPD ($SD_{cal} \div$ SEP). If the validation sample set is derived from the original population, the SD_{ref} of the original reference data (item 16) can be used for RPD calculation.

Items 28–34. These items are all applicable to the evaluations of NIR spectroscopy prediction where the calibrations have been used to analyze samples that have not been used in calibration development. Ideally, these should be fresh samples, but validation sample sets that have been developed using the system described in item 19 can also be evaluated in this way. These statistics are usually reported. The slope (item 31) represents the rate of change in variable Y as a function of variable X. It gives an indication of the potential of the calibration. If the slope differs from 1.00, e.g. by $\pm$ 0.15 (0.85–1.15) or by an even greater deviation, this means that the calibration is likely to be less useful at the extremes. Item 32, the intercept, is optional. It is allied to the slope and can be confused with the bias, with which it is not necessarily associated.

Items 35 and 36. Mahalanobis distances. The inclusion of these two items is optional. The data are not easily available from all NIR spectroscopy software systems, and do not help in the development of practical calibrations, because (a) samples with high GH and NH (WinISI terms related to the Mahalanobis distances) will happen in most sample populations and their variance is an important part of practical calibration and (b) it is possible to obtain high errors in prediction of samples with Mahalanobis distances well within the acceptable range. High Mahalanobis distances indicate that such samples are seriously different from most of the samples and the reasons for the differences should be sought.

Item 37. The SET of the new NIR spectroscopy calibration. This is useful information. The SET (reproducibility) of the NIR method is often superior to the SET of the reference methods. This is because (a) there are fewer sources of error, provided that items five to eight have been carried out correctly and (b) the sample size for NIR spectroscopy is appreciably larger than that of the reference methods, which reduces sub-sampling error. This is particularly applicable to continuous online NIR spectroscopy analysis, because the errors incurred by sampling and cell-loading are eliminated.

Acknowledgement

The constructive comments of Drs. Steve Holroyd and Paul Brimmer are acknowledged with gratitude.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Reference

1. Workman J Jr and Weyer L. *Practical guide to interpretive near-infrared spectroscopy*. Boca Raton, FL: CRC Press, 2008.