PHIL WILLIAMS
with John Antoniszyn & Marena Manley

NEAR INFRARED TECHNOLOGY

GETTING THE BEST OUT OF LIGHT

Near-infrared Technology: Getting the best out of light

Published by AFRICAN SUN MeDIA under the SUN PReSS imprint

This publication was subjected to an independent double-blind peer evaluation by the publisher.

First edition 2019

ISBN 978-1-928480-30-3
ISBN 978-1-928480-31-0 (e-book)
https://doi.org/10.18820/9781928480310

Set in Hind Guntur 10/15
Cover design, typesetting and production by AFRICAN SUN MeDIA

SUN PReSS is a licensed imprint of AFRICAN SUN MeDIA. Scholarly, professional and reference works are published under this imprint in print and electronic format.

This publication can be ordered directly from:
www.store.it.si (e-books)
www.africansunmedia.co.za

NEAR INFRARED TECHNOLOGY

GETTING THE BEST OUT OF LIGHT

PHIL WILLIAMS
with John Antoniszyn & Marena Manley

SUNPROSS

CONTENTS

List of Acronyms Abbreviations

a	Intercept
AACCI	American Association of Cereal Chemists International
ADA	Automated Digital Analyzer
ADF	Acid detergent fibre
AGSA	Australian Grain Science Association
AME	Apparent metabolisable energy
ANN	Artificial neural network
AOAC	Association of Official Analytical Chemists
AOCS	American Oil Chemists' Society
AOTF	Acousto-optic tunable filters
APH	Australia Prime Hard
ASBC	American Society of Brewing Chemists
ASW	Australian Standard White
b	Regression coefficient or Slope
BOD	Biological oxygen demand
CCD	Charge-couple device
CGC	Canadian Grain Commission
CV	Coefficient of variation
CPSR	Canada Prairie Spring (Red)
CPSW	Canada Prairie Spring (White)
CSP	Canadian short process
CSWS	Canada Soft White Spring
CWB	Canadian Wheat Board
CWES	Canada Western Extra Strong
CWRS	Canada Western Red Spring

DM	Dry matter
DON	Deoxynivalenol
ERH	Environmental relative humidity
ESBU	European Society of Brewing Chemists
FN	Falling Number
FT	Fourier transform
FT-IR	Fourier transform infrared
FT-NIR	Fourier transform near-infrared
GH	Global H
GPS	Global Positioning System
GQA	Grain Quality Analyzer
ICC	International Association of Cereal Science and Technology
ICNIRS	International Council for Near-infrared Spectroscopy
IDRC	International Diffuse Reflectance Conference
InGaAs	Indium gallium arsenide
IR	Infrared
LAN	Local Area Network
mid-IR	Mid-infrared
MLR	Multiple linear regression
MPLS	Modified partial least squares
MSC	Multiplicative scatter correction
N	Nitrogen
NDF	Neutral detergent fibre
NH	Neighbourhood H
NIR	Near-infrared
NIR-HSI	Near-infrared hyperspectral imaging
NIR LEDs	NIR light-emitting diodes
NIRS	Near-infrared spectroscopy
NIT	Near-infrared transmittance
NSAS	Near-infrared Spectral Analysis Software
OD	Optical density
P	Phosphorus
PbS	Lead sulphide

PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PLS	Partial least squares
ppm	Parts per million
PSI	Particle size index
r	Correlation coefficient
r^2	Coefficient of determination
RH	Relative humidity
RMSD	Root mean square of the differences
RMSE	Root mean square error
RMSEP	Root mean square error of prediction
RPD	Ratio of (standard error of) Prediction to (standard) Deviation
rpm	Revolutions per minute
SD	Standard deviation
SDF	Soluble dietary fibre
SEC	Standard error of calibration
SECV	Standard error of cross-validation
SEL	Standard error of the laboratory
SEP	Standard error of prediction
SET	Standard error of a single test
S/N	Signal-to-noise ratio
SIMCA	Soft Independent Modeling of Class Analogy
SKSC	Single Kernel Charactersation System
SNV	Standard normal variate
SWS	Soft white spring
TDF	Total dietary fibre
TME	True metabolisable energy
TMR	Total mixed ration
USDA	United States Department of Agriculture
WCWW	Western Canada White Wheat
WRONZ	Wool Research Organisation of New Zealand
WRW	Western Red Winter

Acknowledgements

The authors are indebted to the Cooperative Research Centre (CRC) for Value Added Wheat, North Ryde, New South Wales (NSW), for the sponsorship of the Near-Infrared Spectroscopy (NIRS) course presented in 2003, Vicky Solah from Curtin University, Western Australia (WA), as well as Clare Johnson from the CRC for Value Added Wheat, for the initial encouragement to develop the course. We also acknowledge the valuable comments of Jim Psotka, Philip Clancy, Ian Cowe, Tom Fearn, Diane Malley and Karl Norris with deep gratitude. Stellenbosch University, Stellenbosch, Western Cape, South Africa, is acknowledged for its financial contribution towards publication of this book.

PREAMBLE

Marena Manley
Stellenbosch University
Stellenbosch, Western Cape
South Africa

In Canada, until 1972, milling-quality wheat had always been marketed on the basis of grade determined by hectolitre mass and visual characteristics, which was difficult to verify by purchasers of the wheat. During the late 1960s, the Canadian Wheat Board (CWB) decided to offer the top two grades of Canadian hard red spring wheat at guaranteed protein levels, following a similar decision made in Australia and the USA – the major difference being that wheat marketing was (and still is) under federal government control in Canada. Most of the flour mills operated with, or had access to, laboratories. Overnight protein content became a critical factor in quality assurance. The enormous task of protein content determination became the responsibility of the Kjeldahl laboratory at the Canadian Grain Commission (CGC) in Winnipeg.

Early in February 1970, Dr Phil Williams was appointed as the 'Chemist-in-Charge' of protein-testing operations. Phil accepted this responsibility, and subsequently upheld the 60 years of integrity in quality assurance for Canadian grains. Up until that time the Grain Research Laboratory at the CGC had used the Gunning-Arnold (back-titration) option of the Kjeldahl test. Phil introduced the more convenient Winkler modification even though the test still took about 2 hours per sample. It became clear that an alternative test was needed for testing rail-carloads at the time of unload into grain terminals at shipping ports, where it took only 4 minutes to unload up to 90 tonnes of wheat. The Orange-G dye-binding test was evaluated, but was found to be too variable.

Late in September 1971, Phil heard of a 'miracle' instrument that would do a protein test in 10 seconds, with an accuracy of 0.1%. The instrument was made by the Neotec Corporation somewhere in eastern USA. Phil tracked Neotec to Rockville, Maryland, and phoned the company. The call was relayed to the President of Neotec, Bob Rosenthal, who called Phil within the hour, and described their Instrument. Phil advised his director that they should investigate the instrument, referred to as the 'Grain Quality Analyzer' (GQA).

In November 1971, Phil signed the requisition for what was probably the first near-infrared spectroscopy (NIRS) instrument ever purchased for commercial use in the world. The price was C$7,200. The GQA arrived at the Grain Research laboratory, CGC, on 2 February 1972, a truly memorable day.

Phil had never heard of NIRS until September 1971. He told me that he did not even qualify as a novice. Neither had he heard of Karl Norris, whom he first met in November 1973.

During the next few months Phil regularly communicated with Don Webster, Neotec's Chief Engineer, and just as often with Harold Moyer, Neotec's 'fix-it' engineer. Temperature sensitivity proved to be one of the first challenges in the stability of the instrument. Neotec addressed this issue by controlling the temperature of the lead sulphide (PbS) detectors. Towards the end of 1973, Neotec loaned the CGC three GQA instruments which enabled protein testing also to be done at elevator level. This meant that testing could now be done at the terminals in Thunder Bay and Vancouver. Phil rejected Bob Rosenthal's offer to design an instrument that would do 10 tests per minute because the operators would not be able to load the sample cells fast enough to keep up with the instrument. He suggested that if Neotec could invert the system so that the sample was scanned from below, a tray could be designed that held 10 cells. Neotec agreed and designed the instrument for delivery by 31 March 1974. The Automated Digital Analyzer (ADA) was conceived and a new era, the NIRS era, was born. The ADA enabled the testing of 2500 samples in a seven-and-a-half-hour shift. It was the first near-infrared (NIR) instrument to use a digital computer. This allowed the operator to optimise the wavelengths by changing the 'pulse-points'. It also revealed areas where improvements could be made to bench-top instruments, which had all been driven by analogue computers.

By May 1975, every wheat sample analysed by Kjeldahl, (which became increasingly difficult to run, due to excessive flask breakage) was also analysed by the ADA, and over 200,000 comparisons had been made. The weekly statistics from

all of the terminal elevators looked promising. Overnight, Phil secretly changed the multibillion-dollar protein segregation operation from Kjeldahl to NIRS, using the ADA – as of 1 August 1975, the 60-year integrity of Canada's multibillion-dollar wheat industry was controlled by NIRS. This was the first actual application of NIRS in industry in the world.

From November 1976 to November 1977 Phil got the opportunity to work with Karl Norris in Beltsville, Maryland. This led to a life-long collaboration and beautiful friendship which continued until Karl's recent death at the age of 98. Since Karl's retirement, Phil kept contact with him by phone every few weeks.

In this book, Phil shares nearly 50 years of experience in the application of NIRS, and its implementation in industry. The legacy of his life-long contribution to the application of NIRS is documented and shared. New (and all other) users of NIRS can, for many years to come, benefit from Phil's knowledge and vast experience of the use and implementation of NIRS.

Phil, we salute you for your work and dedication – still to this date. You have ensured that the NIR technology developed by Karl, received its rightful place in industry.

PREFACE

Phil Williams
PDK Projects, Inc.
Nanaimo, British
Columbia, Canada

The driving force behind compiling this book is to provide knowledge on all aspects of near-infrared spectroscopy (NIRS) to potential users, and to users who would like to delve a little deeper into the technology. We have assembled the book to also help in the application of near-infrared (NIR) instruments and technology in the industry. The book has been put together from an NIRS users' course, which was first compiled 16 years ago. The course has been presented in 8 countries, and updated many times based on questions from the audience during the 2-day presentations.

This book covers the essential features for successful NIRS application in a practical and easily understandable format. The content on the chemistry and physics of NIRS is conveyed as such that it can be understood by all users. The importance of accurate reference analysis is emphasised. This is followed by the explanation and discussion of the statistics, software, instrumentation, sampling, sources of variance, calibration procedures and interpretation of calibration statistics based on practical experience. *Suggestions for further reading* are given for those who seek a more in-depth knowledge.

The molecular concept is introduced as a way to envisage the materials for the analysis of which NIRS is to be used. The principle of NIRS analysis differs from classical chemical analysis in that it is based exclusively on the spectra. The spectra are, of course, derived from the chemical and physical make-up of the material, but different growing or processing conditions can affect the molecular arrangement in space and, as a consequence, the

relationship between the spectra and the reported composition that results from the calibration process.

Because of the time and expense involved in the development of calibrations that are applicable with confidence in the day-to-day world of the industry, together with the fact that such work is often proprietary, this type of application does not lend itself to publication in scientific journals. These calibrations call for extensive and expensive research, and the book focuses on the procedures involved in setting up such comprehensive calibration models. Because of the proficiency of modern NIR instruments, the reproducibility and reliability of the future NIRS results will often be superior to those of the reference laboratory.

01

Introduction, history and the economic benefits of near-infrared spectroscopy

The first words

Imagine an analytical technique that uses no chemicals, gives accurate and precise results in minutes or even continuously and is simple to install and safe to use. Near-infrared spectroscopy (NIRS) supplies this dream. It is a rapid method for qualitative and quantitative analysis of a very wide range of materials such as powders, liquids, slurries, solid materials and gases. A 'rapid' method is defined as a method that will provide accurate and precise results in two minutes or less – including the time for sample preparation, as well as presenting the sample and removing it from the instrument. The forté of the NIRS technique lies in its speed. It fulfills the requirement of a rapid method, and is also applicable to continuous on-line analysis. On-line analysis is not only faster, but obviates the need for sampling, and the associated sampling error because everything is being tested. Because of the first-rate precision of the spectral data of modern NIRS instruments, relative to the reproducibility of most laboratories, the NIRS predicted results for any application are virtually more reliable than those of the laboratory.

The concept and principle of NIRS analysis differs from classical chemical analysis in that it is based exclusively on the spectra. The spectra are derived from the

chemical and physical make-up of the material, and can differ among materials with the same apparent chemical composition. For example, consider the spectra of wheat samples from the same variety when grown in the northern wheat-growing areas of Alberta and southern Saskatchewan (Canada). The samples may be reported to contain the same protein and moisture content, but the different growing conditions can affect the molecular arrangement in space and, as a consequence, the relationship between the spectra and the reported composition that results from the calibration process. This is the reason underlying the recommendation for replication of samples of agricultural materials with the same or closely similar reference data. A difference of even 0.1% in protein content represents a vast number of molecules.

Six things are absolutely essential to the successful application of NIRS. These are (1) accurate and (2) reproducible reference analysis, (3) excellent spectral precision, (4) comprehensive sample selection and (5) assembly, as well as (6) efficient sample presentation to the instrument. No matter how comprehensive the chemometric software, it will not fully compensate for imperfections in any one of these prerequisites.

There is a crucial need for education in NIRS technology at a university undergraduate level. This book assumes that the reader knows little or nothing about NIRS. It includes areas of the philosophy of NIRS that do not appear in most texts, but are an important part of the background needed to understand what is involved in an effective application of this fascinating technology. It will not delve deeply into the theoretical aspects of the physics and chemistry of NIRS. There are other books that treat these aspects very well. For those who seek the fundamentals of NIRS, i.e. the physics, the chemistry and the chemometrics, Williams and Norris (2000) (see *Suggestions for further reading*) is recommended. The chapters on these subjects were compiled by the finest of experts.

This book is aimed at teaching the essentials of NIRS to industrial-type analysis. It was originally requested by the Royal Australian Chemistry Institute's Cereal Chemists group (now the Australian Grain Science Association, AGSA). As a result, there are many references to grain and grain products, but the principles are applicable to the wide range of materials and parameters to which NIRS is routinely applied. It is also aimed at teaching the principle of developing calibrations that can reliably be used in the workplace, which is different from simply producing a scientific paper. It is aimed at introducing the reader to the practical aspects of NIRS application that may not be taught elsewhere. It is intended for potential users of NIRS, as well as those who have been operating an instrument for some

time and would like to learn more about the technology. The book addresses all aspects of the practical implementation and use of NIRS. It is also aimed at teaching management why they need NIRS.

The things most clients, including management, really need to know about NIRS instrumentation, are which instrument best suits their operation and how can they use it to improve efficiency and revenue. Operators need to know how to use their instrument, how to ensure that their instrument is working properly and that it continues to do so over time.

Readers of this book are encouraged to read it all. As well as the basics, there are a lot of useful facts and hints buried in the text. It attempts to present the technology in easy-to-understand terms, and includes explanations of words or phrases such as 'stray light' and 'non-linearity' that are widely-used in the technology, but are sometimes not fully understood. Most of the examples are based on applications to the agricultural and food industries. This is because these materials offer the most diversity in terms of physical and chemical composition – the principles explained are however applicable to any type of material. There is an increasing interest in mid-infrared (mid-IR) spectroscopy for some of the applications usually performed by NIRS. The principles of application that this book presents are general, and for the most part can be applied to either near-infrared (NIR), mid-IR or Raman spectroscopy.

There are two main types of NIRS applications. The first type can best be described as the 'feasibility study'. The experimental work involves the assembly of a set of samples, scanning them, determining the reference data, as well as using chemometrics to develop and evaluate the calibration.

The feasibility study can be approached in two ways. The first considers the adoption of NIRS in order to replace an existing method of analysis – it also needs to determine whether NIRS is the most appropriate procedure. Sampling for such investigations are concerned mainly with calibration development. All major sources of variance should be represented. Samples should be assembled for the calibration development of such feasibility studies – representing all of the identified sources of variance and replicated at least three times. The samples should be divided into calibration and validation sample sets. The adoption of NIRS is based on the statistics of the feasibility study and, if these are acceptable, the final adoption of NIRS calls for further work, including a much more comprehensive set of samples.

Another way of approaching the feasibility study is more academic, and generally includes the presentation or submission of a scientific paper. A major portion of this type of study concerns more tenuous applications, e.g. using NIRS as a replacement for a traditional method, such as visual appraisal – an operation which has proven successful over the years. The statistics of such studies are often not sufficiently acceptable to denote practical application, or justify replacement of the time-honoured system by industry. Samples for such feasibility studies are often 'grab' samples, assembled by no particular system. They may be assembled simply by buying them from a store, or stores. The results are published in the form of a scientific paper, thousands of which now constitute the NIRS bibliography.

The second main type of NIRS application constitutes the 'development of calibrations', which is applicable with confidence in the day-to-day world of industry. Because of the time and expense involved, together with the fact that such work is often proprietary, this type of application does not lend itself to publication in scientific journals. These calibrations call for extensive and expensive research, and this book focuses on the procedures involved in setting up such comprehensive calibration models. Samples should encompass all sources of variance, and be replicated several times. The main difference between feasibility studies and calibrations intended for use in industry is that, apart from gross outliers, such as those caused by the incorrect entering of reference data, there can be no outliers in an industrial calibration. This is because every sample that is presented to the instrument in day-to-day analysis has to be scanned and the results reported. The calibration will be based on the reference data, the reproducibility and reliability of which will vary among laboratories. The reproducibility of the future NIRS results will often be superior to those of the reference laboratory.

For operations that involve testing large numbers, such as in plant breeding programmes, or testing of some of the more demanding constituents, or parameters such as Falling Number, calibrations with r^2-values of 0.90 or even 0.85 can be a valuable tool in screening. This type of application will identify samples that are very high or very low in the parameter. Upper and lower limits of acceptability are established by experience. Samples that exceed either limit can be accepted as high or low with confidence. Samples that fall between those limits should be tested by the appropriate laboratory method to determine acceptability into the system, but the screening will eliminate up to two-thirds of the testing, with commensurate savings in time, convenience and expenses.

The technique is **N**ear-**I**nfra**R**ed **S**pectroscopy, and the abbreviation should always include the 'S'. 'NIR' refers only to a wavelength range, i.e. near-infrared and not

to an analytical technique. Alternatively, the word spectroscopy should always accompany the acronym NIR, thus NIR spectroscopy. Since its early adoption as a dependable method for analysis of grain in the early 1970s, NIRS has been incorporated into many types of industry, including materials as diverse as rocket fuel, milk, manure, rubber, petrochemicals, plastics, pharmaceuticals, wood, cosmetics, fish, fresh meats and fruits, confectionery and many more. The most recent explorations have focused on applications to environmental factors, sensory factors, and diagnosis of diseases in animals and people. Many of the suggestions and examples cited in this book are taken from the cereal foods and feeds industries, but the principles are pertinent to application to any type of material.

The most successful industrial applications of NIRS involve prediction of one or more important constituents, such as protein, moisture, sugar or fat/oil contents, for which abundant absorption bands are present in the NIR region. Some functionality characteristics, such as flour water absorption and flour ash can be predicted with confidence by NIRS because these features are strongly affected by constituents, such as protein content, for which absorbers are plentiful. In the case of ash there is no ash in flour until it has been burned, and what is being measured is related to the cellulosic components of the flour, for which absorbers are also clear-cut (Williams et al., 1981). In general, the further an item to be predicted is remote from actual constituents, such as a flavour component, the less trustworthy the application is likely to be, and the less likely it is to be reliable for use in industry. The impact of error on the commercial aspect in industry, and the consequences of erroneous diagnosis in the medical profession are both features that call for careful consideration before the adoption of NIRS as a replacement for existing methods of analysis in either field.

An NIRS calibration translates the spectral data directly into terms of the reference data, whether that be composition, physical or functionality factors. After calibration the instrument multiplies the spectral data from any sample presented to it by the calibration equation which is essentially a constant, so the efficiency of the prediction depends heavily on the quality of the spectra. If the spectral precision is poor this will inevitably result in impaired reproducibility of analytical results.

Having established the need for an NIRS calibration, before attempting any work on NIRS, the **first** priority is to identify the methods used for reference testing and to determine their precision. The results from any laboratory must be accepted as the accuracy from that laboratory. The **second** most important requirement is

to determine the precision of the NIRS instrument for application to the material that is to be analysed (i.e. the spectral precision).

Spectral precision depends on the instrument design (mainly its sample presentation system), the physical nature of the material, the composition with particular respect to moisture content and the skill of the operator in presenting the sample to the instrument. It can be determined by scanning a single sample of the material 10 times, with re-blending and re-loading the sample cell between scans (see Chapter 9, section 9.1.2). After each scan, the scanned sample should not be returned to the original sample. It should be saved in a separate receptacle and returned to the original sample after the determination of the spectral precision exercise has been completed. After the 10th scan, the sample is left in the cell and re-scanned a further 9 times without disturbing the sample in the cell at all.

The standard deviation (SD) of the raw spectral absorption data is calculated for both series of scans at selected wavelengths. Wavelengths 1210 and 2230 nanometre (nm) are suggested. Wavelength 2230 nm is one of the 'classical' reference wavelengths originally used in discrete filter instruments. The SD of the first series represents the overall spectral precision (reproducibility). The SD of the second series represents the contribution of the instrument to the overall spectral precision (repeatability of the instrument). Because there has been no interference caused by the operator, the SD for repeatability is usually very low compared to the reproducibility. No calibration model is needed for this exercise, because it involves only spectral data. An example is given in Chapter 9 (see Table 9.2).

The **third** requirement is to ensure the reliability of the samples used in the calibration development. This involves the identification of all of the sources of variance likely to influence the spectra. The **fourth** requirement is the assembly of a set of samples that includes and replicates all of these sources at least five times. This also involves the sampling technique. The **fifth** prerequisite is the efficient presentation of the sample to the instrument. This is related to the sample presentation system of the instrument, the type of material, and the proficiency of the operator. The type of material refers to its physical form, e.g. solid or granular surface, powder, slurry, fibrous, liquid or high-moisture. The **sixth** and final prerequisite is to establish the accuracy of the NIRS predicted results, relative to the results of the reference analysis, because that is the yardstick by which all future analysis by the NIRS instrument will be compared. Every question that arises from the results of NIRS work comes back to these basic facts, and not even the most sophisticated chemometrics will fully compensate for irregularity in any of these six factors.

1.1 What does the operator really need to know?

This book will explain all of the factors needed to run an NIRS operation. There is much more to setting up and running a successful operation than using the instrument. That is the easiest part. There are things that the operator needs to know, to obtain and sustain the best performance of the instrument, and to obtain confidence in the technology.

Firstly (1), it is important to know what can (and cannot) be measured by NIRS. Secondly (2), one should understand how and where does NIRS fit into the 'bigger picture' of the day-to-day operation. Further items include (not in order of importance) how to:

3. select the most suitable instrument;
4. operate the instrument efficiently;
5. maintain the instrument, to assure continued good performance;
6. monitor the performance of the instrument, including the instrument diagnostics;
7. interpret the results of monitoring;
8. recognise the onset of inferior performance;
9. select samples for calibration;
10. prepare samples for calibration and subsequent analysis;
11. calibrate the instrument;
12. update calibrations;
13. evaluate a calibration;
14. understand the basic statistics;
15. get the best out of the software;
16. extend the scope of the instrument, and to apply it to its best capability;
17. interpret what has happened during development of a calibration; and
18. improve and extend a calibration.

These features will all be addressed in the book, but not necessarily in the above order. Item 2, the 'bigger picture' looks into the future and should be considered at the outset. It refers to the implications of using NIRS in place of the analytical method(s) that have been used hitherto. The three main factors affected are those of the implications on staff requirements, and the obvious need for reduction in laboratory staff which may involve consultations with affected unions. The second factor is the impact of replacing a previous analytical method on the overall economics of the operation. The third factor considers the impact of obtaining essentially instantaneous results for quality control in comparison with the intervals of several hours between sampling, as well as obtaining results by previous methods. Another factor for a commercial analytical laboratory, is

the possibility of introducing significant reductions on the cost of testing for the constituents most frequently tested. This could lead to an increase in clientele and in commensurate revenue. There are several other factors, all of which attest to the benefits of introducing NIRS to many types of operation. Items 4 and 5 are explained in the operating manual that accompanies the instrument at the time of purchase.

Certain parts of the book might overlap, but 'repetition is the corner stone of learning'. More detailed descriptions of areas such as the pure physics and chemistry of NIRS, instrument design and other features are given in *Suggestions for further reading* at the end of each section for those who want to learn more. Background is given and some basics are introduced. The index at the end of the book will help in finding specific areas of the technology.

1.2 What can and cannot be measured with NIRS

The most successful applications are the prediction of composition in terms of constituents such as water, fat, protein, starch, cellulose, and other constituents, all of which have abundant absorbers, such as O-H, C-H and N-H (see Chapter 2). Prediction of functionality factors such as flour water absorption, digestibility, particle size and slurry viscosity carry no direct absorbers. These factors are predictable to a fairly high degree of efficiency because of their physical composition, as well as by association with constituents in their make-up that carry absorbers, in such a way which relates the functionality to the spectral characteristics.

Viscosity and density, the texture of flours and ground grains, as well as solid materials and sludges can be measured by NIRS because they depend to a large extent on particle size, and the way in which the particles arise. Attempts at application of NIRS to factors for which there are no apparent absorbers, such as tastes and flavours in foods, and even the degree of freshness in fruits, vegetables and meat may lead to disappointing results. Constituents that affect these are present in parts per million, such as pesticide and herbicide residues or flavour components. These are difficult to determine with sufficient accuracy to comply with regulations because the absorbers that they contain can be dominated by the same absorbers present in constituents, such as C-H, O-H and N-H, which are present in much higher concentration. It may be possible to determine some such substances in soils, which contain fewer absorbers than most agricultural materials. Heavy metals have no absorbers in the NIR region and are theoretically impossible to predict, but under some circumstances their association with other

materials, such as cellulose, may enable their prediction with a reasonable degree of success (Malley & Williams, 1997).

Differences in the thickness of thin films of materials, such as plastic which may be only a few micrometers in thickness, can be determined using NIRS. The few micrometres represent the thickness of millions of molecules, all of which can be detected by NIRS. When the light energy penetrates the molecular layer as far as the material that it encloses, such as copper wire, the NIRS signal will change abruptly and differences in the thickness of the molecular layer can be measured. What the operator needs to do for the NIRS analysis of any material is to tell the instrument which molecules to identify. This is done by 'marrying' the spectral data to the reference data. The more precise and accurate the reference data, the more precisely the desired information in the spectral data will be identified, and the more effective the NIRS analysis. In the example of the prediction of the thickness of a film, such as the coating of a cable, the calibration can be developed from samples of cables with differences in the coating material and its thickness, and in small differences in thickness that transpire throughout the length of the cable. No chemical analysis is needed.

It is important to determine whether NIRS is a practicable technique to employ in any specific operation. The technique is particularly applicable to situations that call for rapid on-the-spot analysis, such as unloading of farm trucks carrying, e.g. wheat, and also where continuous on-line analysis is required. It is also an important asset to operations that demand analytical methods that involve purchase and disposal of large amounts of chemical reagents – also to operations such as screening or plant-breeding that call for the analysis of large numbers of samples. However, for operations that require only periodical analysis, the economics may favour the traditional 'wet-chemistry' methods.

1.3 The economic benefits of NIRS

Management pays the bills. This book is also designed to help Management understand why their operation needs NIRS, and what are the many benefits that it offers. Probably the most important advantage of NIRS is the speed of the analysis. NIRS provides the user with the same precision (often even better) as any reference method. It does this at a fraction of the cost and in a fraction of the time per test with complete safety and flexibility to take the instrument into the field. The costs of standard 'wet chemistry' tests vary from about C$15,00–800,00 or more per test. An NIRS test costs C$2,00–5,00, but several constituents can be measured at the same time, for the same cost for all the parameters included.

An aspect that Management can accept with confidence is the economic benefit of introducing NIRS to their operation. Diode array NIR instruments are fast enough to allow continuous on-line analysis of material on moving belts and other processing equipment. On-line analysis allows factories to monitor their production continuously which allows them to correct, in minutes, what could possibly become expensive deviations from the quality of their products. Because all of the material is being scanned and analysed continuously, sampling and the associated errors are eliminated.

NIRS has made it possible and practicable for people to test for constituents or parameters for which they had always wanted to test, but were previously not able to, because of restrictions in time and expense. The gains in revenue to all aspects of the grain industry coming from protein as a commodity have been astronomical over the past almost 50 years since protein testing of wheat and barley has become widespread. Other examples include testing for amino acids in grains, fatty acid composition in oils, fibre components, digestibility, energy factors in feeds and forages, and nitrogen and phosphorus in manures, all expensive and time-consuming applications to which NIRS has been successfully applied.

The costs of an NIRS instrument and its calibration must be considered in relation to the intended application. Modern NIR instruments can be expected to function efficiently for 10 years or more. An NIRS instrument that cost thousands of dollars can be used for analysis of commodities or products worth millions. For example, it can truly be said that NIRS has revolutionised wheat handling and marketing. The value of the respect gained in the world of wheat marketing by the Canadian Wheat Industry as a result of their pioneer use of NIRS in 1975, goes beyond pricing. Use of NIRS adds to the integrity, efficiency and flexibility of any operation.

Another aspect of assessing the value of using NIRS is the consequence of not testing, or not testing with sufficient frequency, again due to the time and/or expense of testing. An example here is feed mills that may rely on 'book' values for composition, and as a result may not utilise expensive ingredients, such as protein supplements, with the best efficiency. Analysis of major ingredients, together with analysis of finished products can improve the consistency of performance of the feeds, as well as boost the integrity and reputation of the feed company.

Yet another aspect deals with testing raw materials and developing products during processing in a manufacturing plant. Checking these by reference analysis is expensive, but more important is the time factor. Using 'wet chemistry' methods may take two or three hours, even days, before the plant learns that

their product is deficient in or contains excess of one or more ingredients, or that their products are variable due to inefficient mixing. Raw materials and process control can be monitored instantly by NIR instruments located at key points in the plant. Remote sensing using fibre optics cables and networking software enables several instruments to be controlled from a single computer.

An earlier critical examination of the economics of NIRS demonstrated that an operation involving at least 10 laboratory analyses during each working day would benefit economically by switching from traditional 'wet chemistry' methods to NIRS, while an operation that involved 20 or more daily analyses simply could not afford not to use NIRS.

1.4 Development and implementation history of NIRS

The NIR region of the electromagnetic spectrum extends from about 700 nm to nearly 3000 nm. Because of the limitations of their detectors, most NIR instruments do not record spectra past 2500 nm. There is increasing interest in the mid-IR range. Mid-IR spectroscopists describe wavelength in terms of frequency, recorded as wavenumbers, or reciprocal centimetres (cm^{-1}). The wavelength in nanometres is given by 10,000,000 (10^7) divided by the wavenumber. Similarly, the wavenumber can be determined by 10,000,000 (10^7) divided by the wavelength. The spectral range for mid-IR spectroscopy extends from about 3300–800 cm^{-1} (3000–12,500 nm). The relationship between wavelength and frequency (wavenumber) is non-linear.

First of all, where did NIRS start? Spectroscopists have recognised the existence of absorption of energy that occurred in the infrared (IR) region for many years since William Herschel drew attention to the phenomenon in 1800 (Herschel, 1800). All fundamental absorptions occur in the mid-IR region, and the bands that appear in the NIR region are all overtones or combinations of these (see Chapter 2). Like ripples on a pond, the fundamental absorptions occurring in the mid-IR region extend down to the lower NIRS wavelengths as overtones.

Unlike the sharp spectral peaks seen in the mid-IR region, those in the NIRS region are broad and overlapping. The number of fundamental absorbers that can come from a molecule is given by $(3 \times N) - 6$, where N is the number of atoms in the molecule. Water, for example, with 3 atoms has $(3 \times 3) = 9 - 6 = 3$ possible fundamental absorbers. A molecule such as glucose ($C_6H_{12}O_6$) has a possible 66 fundamental absorbers. All of these fundamental absorbers have up to 4 usable overtones, so that it is easy to see that the NIR region contains many absorbing bands and is

very complicated. Until the introduction of the computer, studies were limited to the fundamental absorbers in the mid-IR region, mainly because absorption bands are sharper in the mid-IR region and there were too many overlapping absorbers in the NIR region. The wealth of absorbers, including overtones and combination bands in the NIR region, is actually an asset to quantitative analysis. The large number of overtones and combination bands offer more flexibility for wavelength selection when absorbers may interfere with one another.

In the late 1950s, Karl H. Norris at the U.S. Department of Agriculture (USDA, Beltsville, Maryland) was the first to consider that this mass of absorbers could be analysed by means of a computer. With a series of co-workers, he demonstrated the feasibility of using the NIR information to determine moisture content of soybeans using 1940 nm as the main absorption band. During his work on the prediction of the moisture content of soybeans, he noted that the measurement was affected by the high protein content of soybeans which can reach over 40%. This was subsequently attributed to a strong protein absorber at 1978 nm. Karl H. Norris is justifiably credited with being the 'father' of modern NIR technology.

Karl Norris (left), the 'father' of NIRS, together with Phil Williams at the 2010 International Diffuse Reflectance Conference (IDRC).

The first commercial application of NIRS in industry is credited to the CGC who adopted it as the method for testing an annual workload of over 500,000 rail-car and cargo samples of CWRS (Canada Western Red Spring) wheat for protein content. Their NIRS operation commenced in early 1972, using the Neotec GQA.

The Neotec GQA Model I used by the CGC in 1972.

The ADA – note the vintage computer (top) and illustration of loading the cells of the ADA sample tray.

The Automated Digital Analyzer (ADA) (Williams, Stevenson & Irvine, 1978) followed the GQA and deserves its place in history The ADA was the first commercial NIRS instrument that employed a digital computer and allowed the operator to optimise the wavelengths by changing the 'pulse-points'. It was designed and built in 1974 by the Neotec Corporation with input from the CGC on sample presentation. Bob Rosenthal was the president of Neotec Corporation at the time, together with their chief engineer, Don Webster. It tested 10 samples of ground wheat per minute, and in 1975 it enabled the CGC to convert their entire protein segregation programme (over 500,000 samples annually) from a 100% Kjeldahl to an NIRS operation – monitored (then) and validated by Kjeldahl testing.

Replacement of the original analogue computer by a digital computer in the ADA (which used the same optical system as the original Neotec 3-filter instruments) also opened the door to solving some of the chronic problems of the early generation of NIR instruments. The ADA operated for 18 years during which it

Bob Rosenthal, a true pioneer and the president of the Neotec Corporation during the early 1970s (top) and Don Webster, their chief engineer.

completed over 11 million tests for protein and moisture content without any serious downtime. Lessons learned from the use of the ADA led to the development of the Neotec GQA 31EL in which Dr Ron Moen, Neotec's statistician, was instrumental.

The NIRS system was in full operation at terminal grain elevators. by late 1976. Conversion from the chemical reference Kjeldahl analysis to NIRS currently saves the CGC an estimated more than C$3 million annually in operating costs besides being an environmentally clean operation. To effect segregation and marketing of wheat on the basis of guaranteed protein content every rail-car load delivered into terminal shipping elevators at the ports is tested and stored on the basis of its protein content. The time of unloading a 90-tonne hopper car is only 3–4 minutes. This type of segregation would be impossible with any other method of testing.

The Neotec GQA Model 31 EL filter instrument. This picture was taken at ICARDA in Syria – probably the oldest, frequently-used instrument in the world in over 30 years.

Switching from Kjeldahl to combustion analysis as a reference method in 1996, the CGC was the first operation in the world to achieve a chemical-free system for large-scale testing of grain for protein content. The use of NIRS in determining the government-guaranteed protein content for their multibillion-dollar wheat industry by the CGC, gave NIRS instant integrity worldwide. The ability of elevator managers to pay farmers on the basis of protein content has generated millions of on-the-spot dollars for western Canadian wheat farmers over the past almost 50 years.

The door had been opened and other groups entered the field. DICKEY-john had already developed the GAC (Grain Analysis Computer) using 6 filters and log 1/R

with no pre-treatment. In 1974, Technicon Industrial Systems became associated with DICKEY-john and David Funk introduced major changes to the GAC. In 1978, DICKEY-john introduced their updated version of the GAC, the InstaLab®. In the meantime, Dr Fred McClure from North Carolina State University, and particularly Dr John S. Shenk from Penn State University had developed computerised spectrophotometers and software for work in the field of NIRS.

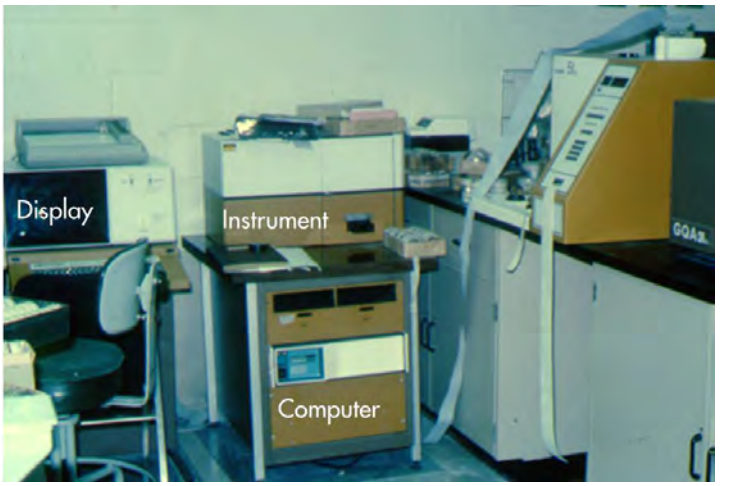
Country primary elevators (top) and a terminal grain elevator.

In 1975–1976, a group of workers in the USA, under the leadership of Dr John Shenk, recognised the potential of NIRS for the determination of quality in forages, the major ingredient in the diets of many domestic animals. The energy and dedication of John Shenk led to the development of the first network of NIR instruments, and the introduction of on-farm testing of fresh and matured forages, using NIR instruments transported in a cargo van. Dr John Shenk, in partnership with Dr F.E. 'Woody' Barton, as well as Rick Hoover and Dr Mark Westerhaus, who later developed WinISI software, developed the first NIRS network. WinISI incorporated many of Dr Shenk's innovative ideas and became the world's most widely-used NIRS software.

The Kjeldahl laboratory at the CGC in 1972.

Between 1974 and 1978 improvements in instrument design and software, together with the perseverance on the application side, expanded applications from protein and moisture to oil, fibre and starch contents. In 1978, Neotec introduced the first computerised NIRS scanning spectrophotometer offered commercially, the Model 6350. This instrument was driven by a digital computer and

operated using a precursor of the DOS-driven Near-infrared Spectral Analysis Software (NSAS), the first software ever dedicated to an NIRS instrument. It operated over a wavelength range of 1100–2500 nm and enabled the use and optimisation of derivatives, as well as the wavelength range. This development extended NIRS testing to constituents such as amino acid and fatty acid composition and parameters such as wheat kernel texture, barley metabolisable energy and others. Since then other companies have developed computerised scanning spectrophotometers and associated software. During the mid-1980s the wavelength range was extended down to 380 nm.



The first computerised scanning NIR spectrophotometer, the Pacific Scientific Model 6350, introduced by Neotec in 1978.

Dr Ed Stark who designed the first diode array instrument in 1991.

Technicon soon followed the Pacific Scientific Model 6350 with their InfraAlyzer Model 500, largely designed by Dr Ed Stark. A new company, LT Industries, was formed by Izaac and Aviva Landa. LT Industries introduced their Quantum series of computerised spectrophotometers. The 1970s was a keynote decade for NIRS.

Probably the last major advance to NIR technology was made in the year 1991 when Ed Stark introduced the diode array instrument that he had designed at the international NIRS conference in Aberdeen, hosted by Ian Murray and Ian Cowe. This was the first ultra-rapid instrument that has since facilitated continuous on-line industrial NIRS analysis of liquids. In 1980, Trebor company introduced the first commercial NIRS

instrument developed for testing whole wheat and barley kernels for protein and moisture content (Williams, Norris & Sobering, 1985). This was a significant milestone since it essentially removed grinding as an important source of error in grain-testing and halved the testing time. It was also an advantage to plant breeders since whole-grain testing allows the breeder to select heritable characteristics in generations as early as F3 and retain the entire sample for planting. The Trebor 90 operated in transmittance mode over the wavelength range of around 950 to 1100 nm using NIRS light-emitting diodes. In 1987 Tecator introduced their Infratec Model 1225. This instrument used a monochromator, and updated versions of the Infratec are widely-used in the grain industry.

The Trebor 90 whole-grain NIRS analyser.

Figure 1.1 shows a typical spectrum of ground wheat from 1100–2500 nm. Note the characteristic 'peak' at 1936 nm. This corresponds to one of the most important absorption bands for water. Other important water bands occur at 964, 1154, 1410 and 1460 nm (these might differ slightly, depending on the instrument).

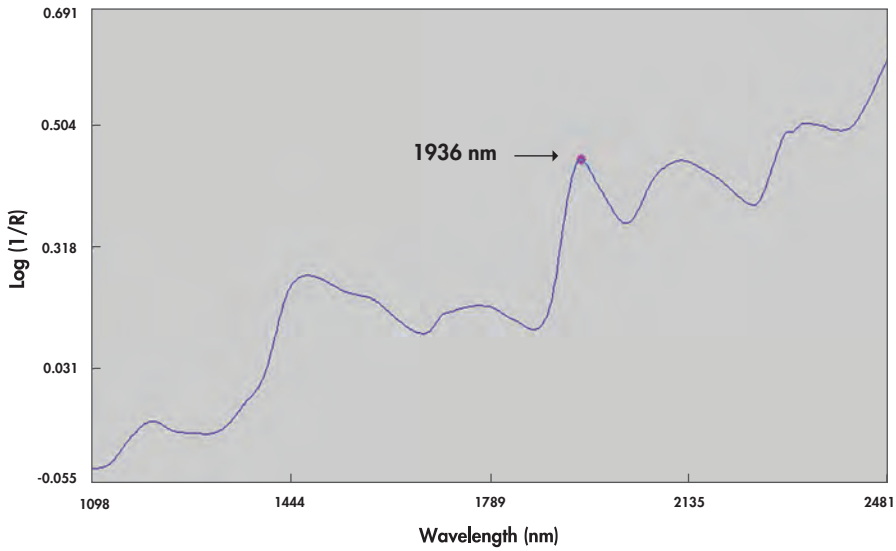


Figure 1.1 Spectrum of ground wheat with the characteristic moisture absorption band indicated at 1936 nm.

Differences in particle characteristics result in differences in the way that light is reflected back from the sample, or through the sample, to the detector. Samples with bigger particle size, including whole grains, cause more energy to be absorbed and less light is reflected. Most NIR instruments record the spectral data as the log 1/R (apparent reflectance), sometimes called Absorbance or Optical Density. Transmittance instruments record spectral data as the log 1/T (apparent transmittance). When a sample is irradiated with light energy, some of the energy is scattered and lost within the sample. Most of the energy is diffusely reflected, some of which is captured by the detectors and used in analysis.

The degree to which the energy is scattered is affected by particle size and shape. Bigger particles, such as whole grains and seeds, cause a higher degree of scatter. This results in more absorbance by the sample and less reflectance. The visible effect is that spectra of larger particles, such as whole wheat kernels, will appear higher in the plot than particles of ground wheat or flour. Figure 1.2 shows log 1/R spectra of ground hard and soft wheats from 400–2500 nm, while Figure 1.3 gives spectra of ground and whole wheat. The main reason for the spectra to rise towards the longer wavelengths is that the absorbers in that region are stronger. The high spectral signals in the visible region, from 400–700 nm show a colour effect. Dark-coloured materials appear higher in the plot than brighter-coloured, such as yellow, materials.

During the 1974–1978 period, the use of first and second derivatives of the log 1/R spectral data was developed and introduced by Karl Norris (Norris & Williams, 1984) (see Chapter 5, section 5.2.5). An excellent description of derivatives is given by Hruschka (2001). Figure 1.4 shows the log 1/R spectrum of a sample of whole-kernel winter wheat, and the 2nd derivative (2 4 4 1) spectrum of the same wheat. The main water band at 1908 nm and its overtones at 1412, 1162 and 982 nm (arrowed) are very clear. Modern NIRS software systems all include the option of using derivatives.

Chemometrics has become a widely-used tool in NIRS. Chemometrics is defined as the application of mathematical or statistical methods to chemical data. The introduction of principal component analysis (PCA) (Cowe & McNicol, 1985) and partial least squares (PLS) regression in the mid-1980s (Martens & Næs, 2001) has resulted in the replacement of multiple linear regression by PLS regression for most calibration development. Subsequently, application of the artificial neural networks (ANN) concept (Borgaard, 2001) has resulted in development of very stable calibrations based on substantial sample sets and has facilitated development of large-scale networking of instruments.

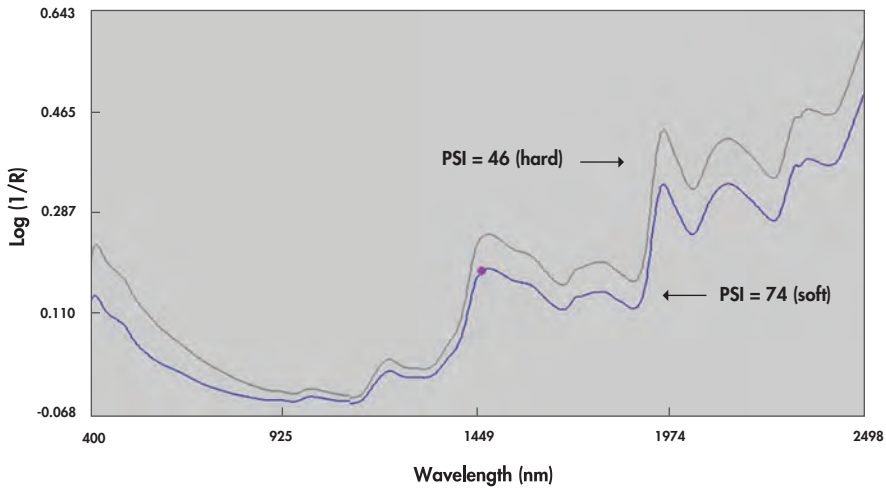


Figure 1.2 NIR spectra of hard and soft ground wheat (PSI = particle size index).

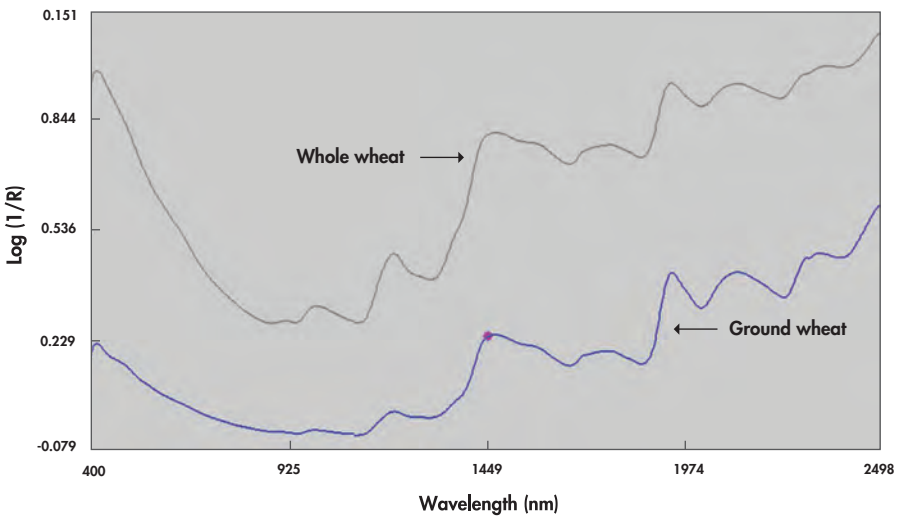


Figure 1.3 NIR spectra of ground and whole wheat.

Advances in computer design and software have kept up with instrument development. These advances have brought NIRS technology to the stage where memory capacity and computing time are no longer obstacles to most operations including storage of large numbers of spectra, calibration development and evaluation, optimisation of wavelength and mathematical treatment, as well as calibration transfer.

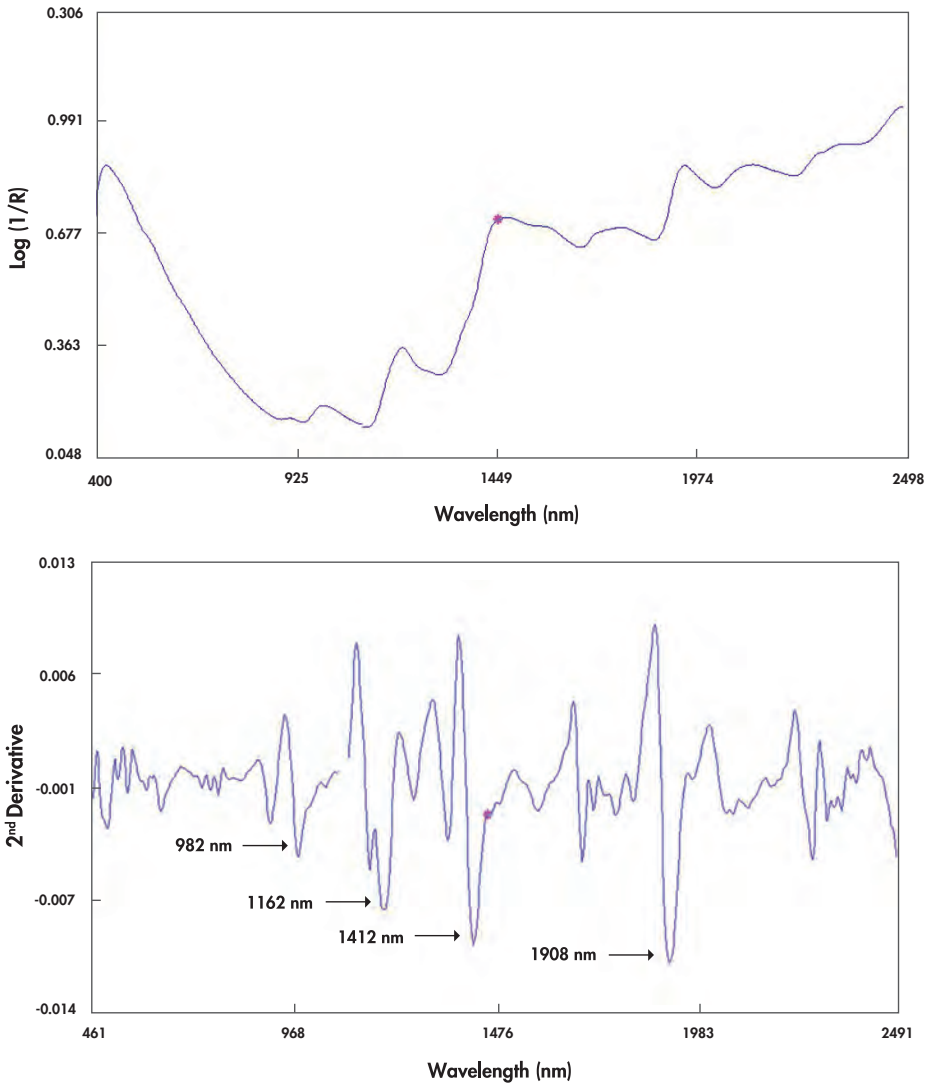
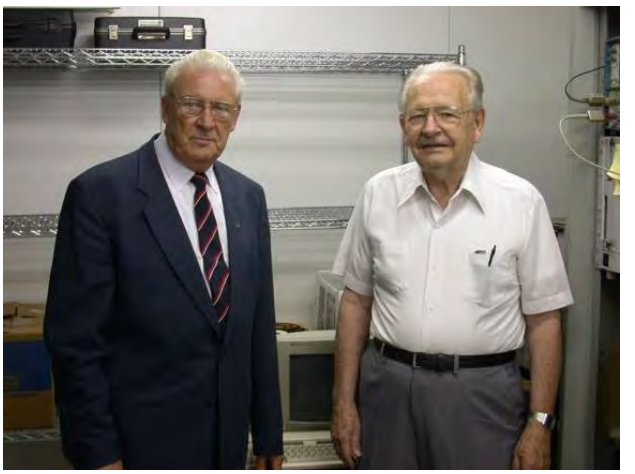


Figure 1.4 Log 1/R spectrum of whole wheat and the 2nd derivative (2 4 4 1) spectrum of the same wheat showing overtones (1412, 1162 and 982 nm) of the moisture band at 1908 nm.

In 1978, Dr Karoly Kaffka was the first purchaser ever of a computerised spectrophotometer, the Neotec, Pacific Scientific Model 6350. It was at a conference organised by him in Budapest, Hungary in 1986 where the International Council for Near-infrared Spectroscopy (ICNIRS) was inaugurated. The meeting was the first of what has been a long series of International biennial NIRS conferences.



Karoly Kaffka (left) and Karl Norris. Two NIRS giants together.

In 1982, Dr Gerry Birth, a physicist working at the USDA laboratory (Athens, Georgia) had masterminded a conference that grew into the International Diffuse Reflectance Conference (IDRC) held every two years in the US. He selected Wilson College in Chambersburg (Pennsylvania, USA) as the venue and organised the conference using the same format as the Gordon conferences; morning and evening sessions with afternoons free for discussions and recreation. The IDRC series quickly became recognised as an excellent forum for discussion of the latest discoveries and revelations of NIRS. It continues to be one of few conferences where beginners and seasoned troupers can intermingle and where newcomers can rub shoulders with the big names and find that they are very approachable.

1.5 Advantages and disadvantages of NIR technology

The advantages of NIR technology:

- Speed – results in seconds, or even continuously, rather than in hours or days.
- Accuracy equivalent to any reference methods.
- Precision (reproducibility) equal or superior to reference methods.
- Low cost per test – low labour costs, no chemicals (to purchase or dispose of).
- Flexibility in sample presentation to the instrument – no other test method offers the flexibility of accepting whole- or ground-state materials in amounts of up to a kilogram of material without affecting the sample in any way.

Dr Gerry Birth who initiated the IDRC series.

- Flexibility in testing – many (up to 12) constituents can be tested simultaneously.
- Environmentally clean – no chemicals.
- Easy and cheap to install – no drainage or exhaust needed.
- Little or no sample preparation.
- Stand-alone instruments, with no peripherals.
- Small instrument size.
- Durability – instruments work well for 10+ years.
- Simple and safe to operate.
- Calibration transferability among instruments of the same model (and even different models).
- Networking – many instruments, even remote from each other, can be networked to use the same calibration, and performance controlled from a single control centre.
- Continuous on-line and interactance analysis (using a fibre optice probe) by NIRS obviate the need for taking samples.

There are also some disadvantages associated with NIR technology. These include:

- Separate calibrations are needed for every commodity and constituent – however, many modern instruments can now be calibrated at the factory and the user simply has to verify performance).
- The need to monitor accuracy and reproducibility (also needed in reference analysis).
- Instruments are expensive to purchase, but quickly pay for themselves in costs saved in labour, installation, time and chemicals.
- Skepticism – many people lack faith in the technology.
- Lack of knowledge as to how to operate instruments most efficiently.
- The crucial need for education in NIR technology at a university undergraduate level.

1.6 Last words on advantages of NIRS

NIRS has revolutionised many aspects of the branches of industry and commerce. Its increasing use is bringing about substantial reductions in the amounts of corrosive and potentially harmful chemicals used in analytical work. The contribution of NIRS to the overall economics of any industrial operation that calls for analytical work is equaled by its contribution to the environment. A sideline effect of NIRS is that it has focused attention of many laboratories on the accuracy

and reproducibility of their 'tried and trusted' reference methods. This is having the effect of causing a gradual improvement in laboratory precision. One of the most important contributions of NIRS is speed of analysis. Time is an important commodity when analysing samples which deteriorate over time. NIRS provides a simple and effective method for rapid analysis of samples that could not really be analysed before NIRS due to the time required. NIRS has opened and will continue to open peoples' minds to a new world of possibilities in analysis.

References

Borgaard, C. 2001. Neural networks in near-infrared spectroscopy. In: P. Williams & K. Norris (eds). *Near-infrared Technology in the Agriculture and Food industries*. 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. pp. 101–107.

Cowe, I.A. & McNicol, J.W. 1985. The use of principal components in the analysis of near-infrared spectra. *Applied Spectroscopy*, 39, 257–266.

Herschel, W. 1800. Experiments on the refrangibility of the invisible rays of the sun. *Philosophical Transactions*, 90, 284–292. https://doi.org/10.1098/rstl.1800.0015

Malley, D.F. & Williams, P.C. 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environmental Science and Technology*, 31, 3461–3467. https://doi.org/10.1021/es970214p

Hruschka, W.R. 2001. Data analysis, wavelength selection methods. In: P. Williams & K. Norris (eds). *Near-infrared Technology in the Agriculture and Food industries*. 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. pp. 39–58.

Martens, H. & Næs, T. 2001. Multivariate calibration by data compression. In: P. Williams & K. Norris (eds). *Near-infrared Technology in the Agriculture and Food industries*. 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. pp. 59–100.

Norris, K.H. & Williams, P.C. 1984. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. Influence of particle size. *Cereal Chemistry*, 61, 158–165.

Williams, P.C., Norris, K.H. & Sobering, D.C. 1985. Determination of protein and moisture in wheat and barley by near-infrared transmission. Journal of Agricultural and. *Food Chemistry*, 33, 239–244. https://doi.org/10.1021/jf00062a021

Williams, P.C., Stevenson, S.G. & Irvine, G.N. 1978. Testing wheat for protein and moisture with the automated digital analyser. *Cereal Chemistry*, 55, 263–279.

Williams, P.C., Thompson, B.N., Wetzel, D.l., McLay, G. & Loewen, D. 1981. Near infrared instruments in flour mill quality control. *Cereal Foods World*, 26, 234–237.

Suggestions for further reading

Burns, D.A. & Ciurczak, E.W. 2007. *Handbook of Near-infrared Analysis*. 3rd Edition. New York, USA: Marcel Dekker, Inc.

Osborne, B.G., Fearn, T. & Hindle, P.H. 1993. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. 2nd Edition. Harlow, Essex, United Kingdom: Longman Scientific and Technical.

Roberts, C.A., Workman Jr., J. & Reeves III, J.B. 2004. *Near-Infrared Spectroscopy in Agriculture*. Madison, Wisconson, USA: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.

Workman Jr., J. & Weyer, L. 2008. *Practical Guide and Spectral Atlas for Interpretive Near-infrared Spectroscopy*. 2nd Edition. Boca Raton, Louisiana, USA: CRC Press, Taylor and Francis Group.

Williams, P.C. & Norris, K.H. 2001. *Near-infrared Technology in the Agriculture and Food Industries*. 2nd Edition, St. Paul, MN, USA: American Association of Cereal Chemists.

02

Basic physics and chemistry of NIRS

The efficiency of NIRS analysis depends heavily on the integrity of the reference analysis and the quality of the spectra. However, NIRS works because of the physics and chemistry of all of the materials to which it is applied. The physics affects the technique by its influence in the way that the energy is reflected from the sample. This is a combination of the size, shape and general physical make-up of the material (solid, particulate, slurry or liquid). A degree of scatter is introduced as a result of interactions among these physical factors and the sample preparation process. The degree of scatter is the degree to which the energy with which the sample or material is irradiated, is diffusely reflected back to the detectors. The chemistry affects the technique by the arrangement of the atoms of the molecules. It is the vibrations of the molecules and groupings of atoms of which the materials are composed that are measured by NIRS, which is then translated into the constituents and functional parameters that its users need.

This chapter covers the essentials of the physics and chemistry. It will not delve into the laws of physics or the various molecular vibrations and laws of chemistry as they apply to NIRS. Full details are cogently explained in the *Suggestions for further reading* towards the end of the chapter for those who want to learn more.

2.1 The physics

NIRS is a 'marriage' between spectroscopy and applied statistics. The technology is derived from the basic chemistry of the material to be analysed, but is strongly influenced by some laws of physics. Of these, the most important is light scattering. Our ability to see objects depends entirely on detecting the light energy diffusely reflected from them. The light is reflected at different wavelengths of the electromagnetic spectrum, including the visible range from about 380 nm at the violet end of the spectrum to about 780 nm at the red end. Because of this, the object appears in different colours to the human eye. The human eye can detect several thousand different shades of the primary colours. Because we have two eyes we are able to view the object in the three dimensions – length, width and depth. Because of the ability of light energy to penetrate into the cells that make up a material, and even into the molecules, NIRS can be regarded as having super three-dimensional capability.

There are two types of reflectance. Surface reflectance is termed 'specular' reflectance and does not carry any information about the object. Diffuse reflectance comes from within the object and carries information concerning its composition and other characteristics. Imagine that you are outside a building. Light is being reflected from the surface of the windows at an angle. This is surface or specular reflectance, it carries no information and you cannot see into the room. If you now approach the window and look through it directly, you will be able to see everything inside the room within your vision since you are able to detect the light that is being reflected from all of the objects in the room. This is similar to the diffuse reflectance or transmittance that carries information on composition and functionality from within the sample to the detector of an NIR instrument. In terms of NIRS, diffuse reflectance from a powdered or granular sample is a type of multiple reflectance of the incident light diffusely reflected from the surfaces of all of the particles or granules (including whole grains or seeds), and even from the molecules of solid materials and liquids that make up the sample as it is scanned. The fraction of the incident light energy that exits the sample and reaches the detectors, is the energy used in NIRS analysis.

Some workers prefer to use 'reflection' or 'remission' rather than 'reflectance', and 'transmission' as opposed to 'transmittance' when discussing NIR technology or instruments. Reflection means the act of reflecting, which is the change in direction of energy at an interface between two different media, such as air and glass or any type of material. In NIRS, the energy that carries information is reflected from within the sample by diffuse reflectance. Because by definition reflection

occurs at the surface, it is specular reflection and contains no information about the material, so there is little value to the NIR spectroscopist from reflection. Remission means reduction or diminution and is not appropriate for use in NIR technology.

Reflectance means the ratio of the amount of energy (in NIRS this is energy from light) reflected from the surface to the amount of energy that originally strikes the surface. This is what is actually measured by an NIR instrument. Transmittance is defined as the fraction of energy that passes through a sample, relative to the amount of energy that strikes the sample. This is what is measured by near-infrared transmittance (NIT) instruments. In terms of spectroscopy, transmission is not clearly defined. As such, for practical purposes, reflectance and transmittance are the preferred terms in this book.

The different pathways that light energy can take through or from a sample are illustrated in Figure 2.1. Some light energy (1) is reflected directly from the surface and carries no information (specular reflectance). Some of the energy (5) gets completely scattered within the sample to the extent that it never emerges. Some energy (4) passes right through the sample in transmittance mode to reach the detector and is used in NIR transmittance spectroscopy while other energy is diffusely reflected (3) or transmitted (6) from within or through the sample, but does not reach the detector. The energy that is used by NIRS (2 and 4) is diffusely reflected from within, or transmitted through the sample, and reaches the detector(s).

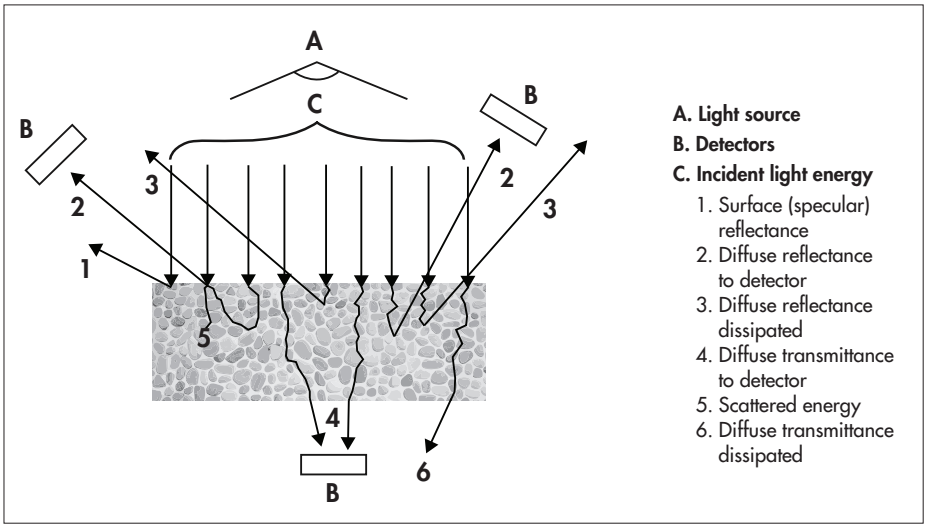


Figure 2.1 Pathways of light through a sample.

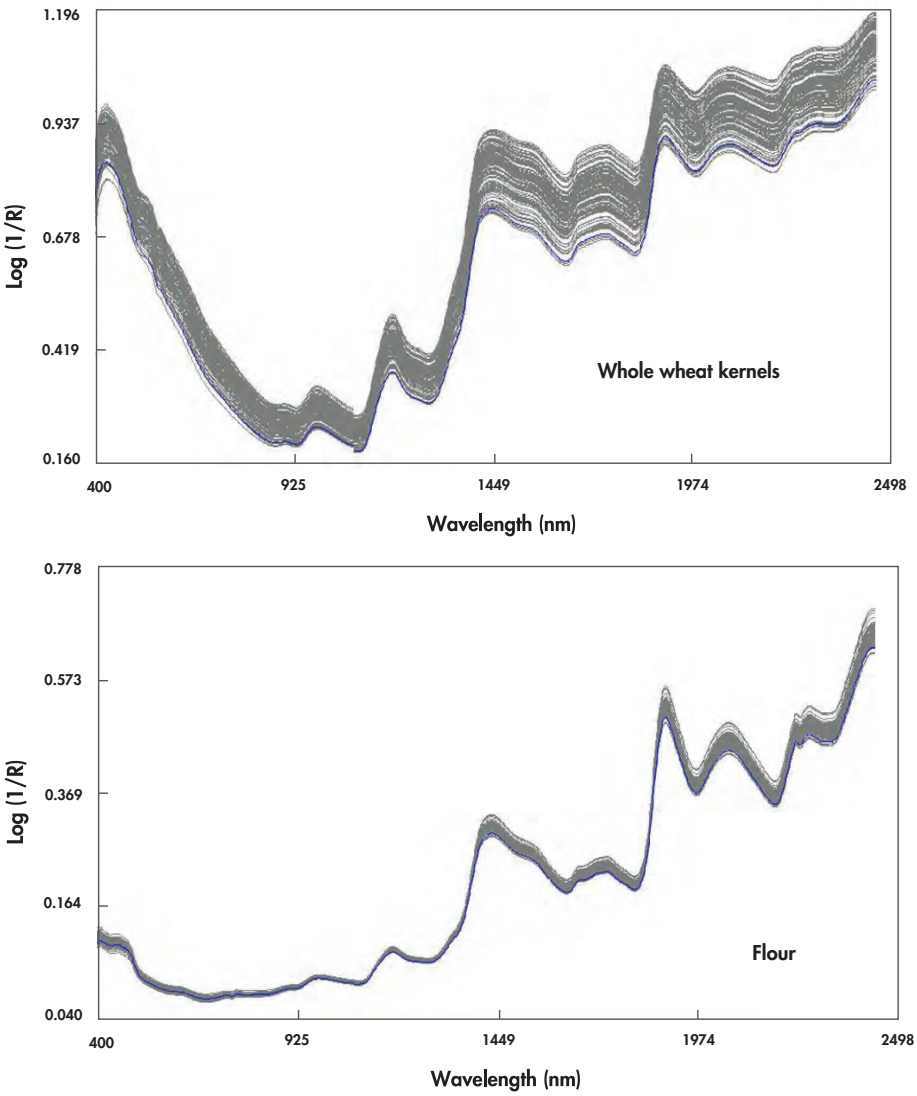


Figure 2.2 Spectra of hard red spring whole kernels and flour, milled from the same wheat samples.

Diffuse reflectance from solid materials, such as wheat flour or whole seeds of any type, is affected by the physical characteristics of the material, including particle size and shape. Figure 2.2 shows the spectra of about 50 samples (whole kernels) of western Canadian hard wheat. All of these are classified as hard, on the basis of the PSI test (Approved Method No. 55-30.01; AACC, 2000a). The spectra show considerable variation on the y-axis (log 1/R or Absorbance) related to variability in size of the kernels and to differences in composition. This spread in spectra throughout the spectrum is what is often referred to as 'scatter'. Note the very large spread of spectral data above 1450 nm. The main reason underlying why

the scatter appears to increase at the higher wavelengths is that in this region the absorbers are stronger and the scattering effect is more pronounced. Whole kernels of wheat or other grains can be considered as particles, even though these particles are much bigger than those of ground grains.

Figure 2.2 also shows the spectra of flours milled from the same wheat samples. The spread (y-axis) or scatter in spectral data is smaller. This is due to the smaller particle size and greater uniformity in particle characteristics in flours as compared to wheat kernels. The flour spectra show much less variability in particle size and less variance in spectral data. Partly because of the lower degree of scatter, the spectral data afford excellent prediction of composition and functionality factors. The principles of scatter are explained lucidly by Birth and Hecht (2001) and Dahm and Dahm (2007), see *Suggestions for further reading.*

2.2 Texture

Most solid biological materials are cellular in structure. Non-biological materials, such as the many forms of plastics are not cellular in the same way, but they are all composed of molecules. An NIR spectrophotometer sees past the surface of the material being scanned, through the cells or matrix of which the material is composed, and into the actual molecules of which the material is made. The irradiating energy that carries information to the detector and computer penetrates not only the cellular structure of the material that is being analysed, but also the molecules that make up the cells. As a result, the pattern of energy that is diffusely reflected from within the sample is affected by the internal cellular structure or architecture of the material.

The architecture of the material is the way in which the cells or particles are combined to create the material. Differences in the association between this architecture and the density of the material create differences in texture. Factors that cause changes in the texture of solids, grains or slurries can be measured by NIR instruments using the diffuse reflectance from or through the sample. Prediction of viscosity in slurries is another application that comes from this concept. In Chapter 7 (see Figure 7.3) differences in the number of particles in a slurry are shown which can affect the diffuse reflectance. When less particles are present, the samples should be scanned in transflectance mode (see Chapter 7, section 7.7.2). In the presence of more particles, more signal is returned to the detectors and the system becomes progressively denser until it is a true diffuse reflectance system.

One of the most important of these factors, causing changes in texture, is the degree of hardness or softness in cereal grains such as wheat. Kernel hardness or texture can be measured by the PSI method (Approved Method No. 55-30.01; AACC, 2000a). The PSI test is a grinding/sieving test with a high degree of precision. The cellular structure of hard wheat kernels is radically different from that of soft wheat kernels; this characteristic is strongly controlled by genetics (Symes, 1965). Because of these differences in the architecture of the kernels, the particle size of ground hard wheats is larger than that of soft wheats to the extent that wheat kernel texture can be reliably predicted by NIRS (Approved Method 39-70.02; AACC, 2000b). The cell walls of hard wheat are thicker than are those of soft wheat to the extent that soft wheat cells may rupture as the cells fill with starch during maturation. These differences in the cellular structure of the kernels also cause differences in the path of light energy through intact cellular kernels. This enables wheat kernel texture to be predicted in whole wheat kernels by NIRS.

2.3 The chemistry

The tendency for users of NIRS is to think of the composition, particularly in terms of the reference data, for the prediction of which the technique is to be used. The concept and principle of NIRS analysis differs from classical chemical analysis in that it is based exclusively on the spectra. The spectra are, of course, derived from the chemical and physical make-up of the material, but the spectra can differ among materials that are reported as having the same chemical composition. For example, consider the spectra of two samples of the same variety and genetic constitution when grown in the northern wheat-growing areas of Alberta and southern Saskatchewan, Canada. The samples may be reported to contain the same protein and moisture content, but the different growing conditions can affect the molecular arrangement in space and, as a consequence, the relationship between the spectra and the reported chemical composition that results from the calibration process. Turning to materials such as silages, the materials will be composed of a range of plant species and substances of different spectral composition, but of similar chemical make-up.

The molecular concept of material is introduced here. To begin with, the hydrogen atom can be visualised as a basketball (the nucleus) with a pea (the electron) circulating it rapidly and at a considerable distance from the nucleus. The unpredictable pathway followed by the electron is called an orbital. The atoms of all elements have similar basic construction, so continuing this concept, no matter how solid a material may appear, it consists mainly of space. The degree of solidity, that differentiates between gases, liquids, slurries and solids is a function

of the density with which the molecules are combined in the material, which is also affected by the electronic structure and density of the atoms themselves.

Molecules, or groups of molecules that are capable of reaction as a group, continuously react with each other. One of the most functional of these is the water molecule. The water molecule contains two hydrogen atoms and one oxygen atom. Oxygen has two pairs of free electrons and two water molecules can form up to four hydrogen bonds. Hydrogen bonds can link water molecules together or link water molecules to other molecules, such as protein or oil. They can form links between molecules of these and other constituents which affect the functionality of the materials. This interaction results in changes in all materials, many of which can be detected and measured using an NIR instrument. Hydrogen bonds are actually coordinate covalent bonds. These are similar to covalent bonds, but not as strong. Nevertheless, because there are so many of them they are very effective. Ice is formed completely from hydrogen bond-linked water molecules, and a metre of ice can support a heavy load.

Avogadro's number gives the number of molecules in a gram-molecule as 6.022×10^{23}. A gram-molecule is the molecular weight of the molecule in grams, so a gram-molecule of water would weigh 18 g. To give some idea of the significance of this, assume an arbitrary value of 250,000 to be the molecular weight of a (small) wheat protein, 1,000,000 to be that of starch and the value of 932 to be the molecular weight of a typical triglyceride. Then a single kernel of hard red spring wheat, weighing about 35 mg, with 12% protein, 12% moisture, about 65% starch and 2% oil would contain approximately 1.4×10^{20} molecules of water, 2.5×10^{17} molecules of protein, 4.5×10^{18} molecules of oil and 1.43×10^{17} molecules of starch – plenty of molecules to react with light, all moving around constantly. A consequence of this is that an error of as little as 0.1% water, for example, represents billions of molecules – all with the potential for hydrogen-bonding. The molecular composition of the sample will not change, but the spectral composition can change as a result of the different degrees of hydrogen bonding.

Hydrogen bonding can cause associations between constituents, such as protein and oil. Such associations change the shapes of the molecules and consequently the spectral features, as well as the functionality. Even when chemical analysis indicates that the composition of two samples of a material such as wheat are identical, with differences that are usually reported to only 2 decimal places, the samples may actually differ in the 3rd or 4th decimal place. Such differences result in differences in the order of billions of molecules undetectable by reference chemistry, but spectrally detectable by spectrophotometers.

Metals ranging in atomic mass from magnesium (24) to lead (207) become associated, but not necessarily in direct combination with proteins and cellulose in agricultural materials, and affect the shape of the molecules. The presence of metals such as calcium, magnesium, potassium and manganese, when absorbed at different levels in soils, can affect the spectra of crops. This is why replication is strongly recommended when selecting samples of agricultural materials with similar composition for calibration development. These differences in spectral composition, incurred by the differences in molecular composition, confer stability to the calibration. Differences in spectral characteristics of similar types of material of the same reported chemical composition offer a possible explanation for the occurrence of what are regarded as outliers. When NIR spectrophotometers are used in industry, the calibration equation is regarded as a constant and directly applied to the scanned sample. Spectral precisison is usually superior to that of laboratory analysis. Provided that the samples assembled for calibration development have included all of the identifiable sources of variance for commodities such as grains and seeds, the NIRS predicted results can be expected to be more reliable than those of the laboratories from which the reference results originate. Referring to wheat protein content, differences of up to even 1% can be caused by differences in the spectra that have come about as a result of external influences that have affected the spectra but not the composition, as determined by chemical analysis.

The position of the atoms in the molecules influences the shape of the molecules in space, as well as their reactivity. The presence of a particularly reactive substance even at the parts per million (ppm) level, would be present in the form of billions of molecules. This could influence the shapes of bigger molecules and affect both their behaviour and spectral signals. The chemical constitution of some constituents, such as cellulose, causes them to exist in plant materials in the form of characteristic shapes – some of which are microscopically visible. The shapes of large molecules such as starch, cellulose and proteins can be expected to change to a lesser degree than the shapes of smaller molecules, as a result of activities such as hydrogen bonding. The cellular structure can influence the pattern of light passing through the cells. An NIR instrument can be imagined as being a super microscope that sees through the surface, through the cells that form the materials, as well as through all of the molecules that form the cells.

2.4 How does NIRS work?

It all comes down to chemistry. All agricultural and food materials are built from cells that are composed of molecules, and contain other molecules. These

molecules consist of atoms and groups of atoms linked together in various combinations by various types of bonds. Practically all organic materials consist of molecules, the atoms of which are bound by covalent bonds. Covalent bonds differ from ionic bonds. In covalent bonds, electrons are *shared* between adjacent atoms in a molecule. In ionic bonds, electron transfer occurs between the atoms. In inorganic materials, including common acids and bases, and salts such as sodium chloride in solution in water, the atoms form ions which are linked by ionic bonds.

2.4.1 *Vibration of molecules*

All molecules are continually vibrating at specific frequencies at what is called the ground level. Irradiation of materials by an energy source such as light causes some molecules to change their vibrations from one energy level, e.g. the ground level, to another. These transitions are constantly taking place, and when they happen energy is absorbed. Such absorption occurs when the vibrations at a certain frequency coincide with those of a molecular grouping in the material being scanned. These transitions are explained in more detail in Dahm and Dahm (2007), as well as Colthup, Daly and Wiberley (1975), see *Suggestions for further reading*. Because the fundamental absorbers and their overtones happen over more than a single nanometre, they are called absorption 'bands'. The absorptions of energy can be detected and measured spectroscopically. These are called the fundamental absorbers. They all occur in the mid-IR region, mainly between 3000 and 12,500 nm, and become repeated in the NIR region as overtones (the 'ripples'). Some overtones also occur in the area between 1800–2500 nm, but most of the overtones appear at shorter wavelengths.

2.4.2 *Functional groups, combination bands and overtones*

Certain groups of small atoms, such as carbon-hydrogen (C-H), oxygen-hydrogen (O-H) and nitrogen-hydrogen (N-H) that are parts of molecules, absorb at characteristic wavelengths. These are called functional groups, an example of which is the -CH_2 (methylene) group. Most of the major constituents of agricultural and food materials contain C-H and O-H groups, while proteins contain all three, together with other formations, such as carbon-nitrogen (C-N), carbon-oxygen (C-O), sulphur-hydrogen (S-H) and phosphorus-hydrogen (P-H).

Absorptions at different areas of the wavelength range between 700 and 2500 nm are all overtones or combination bands which can be assigned to these individual functional groups. Sometimes two absorbers coincide to the extent that an absorption band appears near the sum of the frequencies of the two (sometimes 3) fundamental bands. These are called combination bands, and

combinations of more than one fundamental vibration appear in the NIR region. These combinations, and their overtones, can be found in the wavelength areas between about 800–1100, 1350–1550 and 1850–2500 nm.

Imagine a round pool of water with a perfectly flat surface. Now throw a pebble into the center of the pool. The big splash is the fundamental absorber, and is much bigger than any of the ripples. The ripples that spread outwards are the overtones. The first ripple is the biggest and analogous to the first overtone. Theoretically there are an infinite number of overtones, but like the ripples themselves, the further away that they occur from the fundamental, the smaller and weaker do they become. For practical purposes, only the first four overtones are sufficiently strong to be of much use in NIRS analysis mainly because of interference from stronger overtones. Now imagine throwing two pebbles into the pool. They will fall at slightly different positions, but will cause two splashes which will overlap each other closely. This is comparable to the creation of combination bands. Both splashes will cause ripples and overtones. Because of the closeness of the two splashes that form the combination, their ripples and overtones coincide with each other. The appearance of colour combinations, such as orange and turquoise, can be compared to combinations in NIR technology.

Combination bands and their overtones occur in the NIR region, between as low as 800, and up to 2500 nm. Although some overtone and combination bands appear to be higher in intensity than others, some of which occur in the same area, they may not be selected in development of a calibration model – partly because they interfere with one another when the spectral data are regressed against the reference data. When this happens, the highest overall correlations between the spectral and reference data may include absorbers at wavelengths of lower intensity. The use of these absorbers is another reason for including many samples in the development of calibration models. The smaller overtones may not be prominent in some samples (of the same material), and replication of spectral variance is necessary to give stability to the calibration model.

It is interesting, and to a certain extent informative to be able to relate absorbers selected during calibration development to 'classical' absorbers that have been documented by spectroscopists (Colthup et al., 1975; Workman & Weyer, 2012 see *Suggestions for further reading*). Because of the occurrence of up to four 'useful' overtones and combination bands, more than one of these groups is absorbing at any wavelength point, which complicates assigning wavelengths selected during calibration development to specific absorbers. Absorptions for individual atomic groups, such as -CH_2 have been identified over a range of wavelengths that can

overlap the ranges at which other groups absorb. If a particular wavelength is selected for calibration development, the actual group or groupings responsible for the absorption selected by the computer for use in development of the calibration model, may not be obvious. Most of the early work on identification of absorbers was done using pure organic compounds. These are relatively uncomplicated and easier to work with than, e.g. maize silage, where absorbers from sugars, starch, cellulose, water, protein and oil all interact with one another. These interactions complicate the development of calibration models for prediction of composition of these materials.

It is the interaction among all of the absorbers, large and small, that establishes the overall spectral variance. At any wavelength point one or more absorbers will predominate, and their influence will extend over wavelength points adjacent to that wavelength. PCA identifies the areas of wavelength that account for the variance and categorises them on the basis of the proportion of the overall variance that is explained. The first principal component (PC) may explain >80% of the variance, but it is the interaction among all of the absorbers active in that wavelength area that is combined to give the first principal component. The same process is applied to the absorbers in the spectral variance remaining (the residual variance), until most of the variance has been explained. If the total amount of variance explained falls far short of the total variance, e.g. by 15% or more, the application of NIRS to that analysis is questionable for practical purposes since the unexplained spectral variance is not related to the composition of the material.

When the PCs are displayed, the 'peaks' show the actual wavelengths where the combined intensity of the absorbers has exerted the most influence – whether or not the spectral data have been pretreated. These interactions become most important when NIRS is applied to the prediction of parameters other than discrete constituents such as moisture, oil or protein content.

PLS regression adds variance contributed by the reference data to the overall variance, but the principle of selection of the most important wavelength areas is the same. Display of the PLS factors may show 'peaks' at slightly different wavelength areas than display of the PCs. This shows the influence of the reference data on the overall variance.

2.4.3 Calibration process

The calibration process tells the computer which absorbers to select to correspond with the reference data. The calibration models instruct the instrument as to which wavelengths to monitor in order to provide the information needed by

the operator. Substances such as proteins contain many N-H groups, fats and oils contain many C-H groups, water contains O-H groups, and so on. By making measurements at the wavelengths at which these groups are known to absorb energy, it is possible to determine the amount of protein, moisture, and other constituents present. It has become accepted among NIRS users that the $-CH_2$ wavelength area is assigned to oil (and fat) and the O-H area to water, but it is important to remember that $-CH_2$ and O-H are present in constituents other than oil and water. Protein, starch, oil and cellulose all contain $-CH_2$ groups and O-H groups. Merely monitoring wavelengths where C-H, O-H and N-H are known to absorb is not sufficient to assign absorbances at wavelengths characteristic of these known absorbers to the specific absorbers because some of them overlap each other.

Another factor that complicates assigning wavelengths selected during calibration development to specific absorbers, is the effect of interactions among constituents on the actual spectra of the constituents. Table 2.1 and Figure 2.3 show how the positions of the main absorption bands of water can change, depending on the material that contains the water. The data of Table 2.1 were recorded as the 2nd derivative of the two main water bands (segment and gap each of 4 wavelength points). Figure 2.3 shows how the positions of the water bands change in different materials. In the cases of low moisture maize, wheat flour and starch it is possible

Table 2.1 Changes in position of water absorption bands

Material	Position of absorption bands		
	1900 nm area	1400 nm area	1150 nm area
HM maize*	1894	1408	1154
LM maize*	1912	1426	1160
Hog manure	1892	1414	1152
Starch	1928	1430	1158
Flour	1928	1430	1158
Fresh chicken	-**	1396	1150
Hemp seed	1908	1414	1162
Clay mineral	1910	1410	1152
Oats	1912	1428	1160
Water	1904	1410	1154
Range	36 nm	34 nm	10 nm

* HM = high moisture; LM = low moisture
** Because of the very high moisture content of fresh chicken the 1900 nm water band is indistinct

that the bands at 1426 and 1430 nm are the result of O-H groups on the glucose molecules of the starch or cellulose, rather than on actual water. The curious small band in the 1150 nm area is included. It is not assigned to any functional group in the literature, but it occurs in many materials and is more stable than the bands in the 1400 and 1900 nm areas.

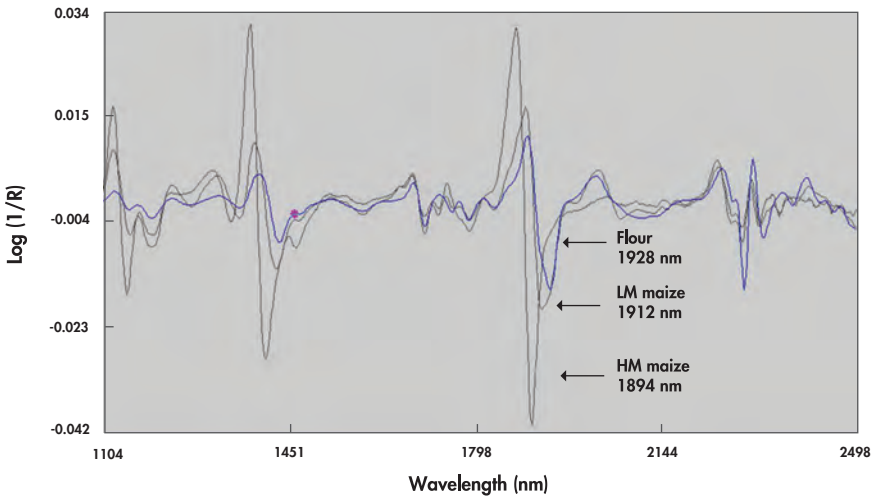


Figure 2.3 Changes in position of water absorption bands (LM = low moisture; HM = high moisture).

The 'undulating' appearance of NIR spectra come about because the absorbers overlap each other, which hides the sharpness of the absorbers. The actual positions of the absorbers can be seen by developing the 2nd derivative of the log 1/R spectra (see Chapter 1, section 1.4 and Chapter 5, section 5.2.5). By making preparations of starch, protein, cellulose and oil as free from other constituents as possible, scanning several samples of water and these constituents several times over the full NIRS wavelength range over a period of time, and then averaging the spectra, it has been possible to identify the wavelength bands associated with the main constituents of agricultural and food materials. Table 2.2 summarises the most important bands for common constituents in biological material. Because of minute differences among instruments, the exact position of these bands may not be the same when materials are scanned using different NIR instruments, and may differ by 1–5 nm. Also, the positions of the absorption bands in different materials are affected by interactions among the constituents (Table 2.1, Figure 2.3), as well as by temperature.

Due to interactions among the 1st and 2nd overtones of some constituents, the small bands (3rd and even 4th overtones) are sometimes selected by the software

in developing calibrations, provided they are consistent in position. Furthermore, some constituents are highly correlated with one another, and the software may use a wavelength associated with one constituent in development of a model to predict another constituent with which it is correlated. For example, prediction of protein content in canola or soybean may use a wavelength associated with oil, because there is a strong negative relationship (r = −0.80 to −0.90) between the oil and protein contents of these materials.

Any of these, even the small bands, may be used in calibration model development. The relative degree to which the bands appear in the spectra of different agricultural materials may also change. For example, the very small and often indistinct water band at 2204–2208 nm is prominent in the spectra of soil (see Chapter 11, section 11.17) and also occurs in the PLS factors used in development of calibration models for several constituents, but it is not prominent in the spectra of wheat or other grains. Because of its prominence in the spectra of soil, it is believed to be associated with silicates.

Table 2.2 Wavelengths of principal absorption bands for common constituents of biological material

Wavelength (nm)					
Strong	Fair	Weak	Strong	Fair	Weak
Protein			Oil		
		708*			660*
		808*		758*	816*
		868*		928*	1210
	908*	982*		1042*	2144
	1018*	1140	1724	1162	
1692	1186	1276	1762	1390	
1734	1428	1360	2306	1410	
1930	1498	1454	2346	1896	
1978	2202	1578		1932	
2054	2308	1628	1724	2008	
2172	2346	1798	1762	2120	
2274		1824	2306	2270	
2466		2108	2346	2384	
		2380			
		2418			

Table 2.2 Wavelengths of principal absorption bands for common constituents of biological material (continued)

Wavelength (nm)					
Strong	Fair	Weak	Strong	Fair	Weak
Cellulose (dry)			Water		
		678*	1906	1410	758*
	758*	816*		1460	964*
	982*	914*		2486	1044*
	1216	1004*			1154
1364	1488	1156			1342
1428	1588	1272			1778
2054	1702	1636			2208
2076	1828	1768			2384
2104	1918	2188			
2270					
2332					
2480					
Starch			Starch (dry)		
		614*		610*	
		758*		754*	820*
	914*	1160		914*	1062*
	986*	1268		980*	1262*
	1200	1360		1204	1272
1432	1584	1750		1360	1574
1928	1700	1826	1432	1700	1750
2094	1780	2188	2282	2044	1784
2282	2370		2322	2094	1910
2318	2474		2446	2478	2180

*These bands are probably 3rd and 4th overtones, but are relatively strong absorbers. The positions of the bands may differ slightly among different spectrophotometers, even of the same model.

This leads to another key point. Prediction of composition depends on the fact that the constituents, such as water, fat and protein all have characteristic absorbers, arising from C-H, O-H stretching and deformation. However, functionality factors such as kernel texture (degree of hardness or softness), metabolisable energy or slurry viscosity and density are not directly associated with specific absorbers. Their prediction depends on the association between their constituents and the

degree to which interactions by these constituents result in the functionality. Because the prediction of functionality is really a secondary issue, and the influence of absorbance at the wavelength points of the constituents are twice-removed, the efficiency of the prediction of functionality factors may not be of the same standard as that of composition factors.

Attempts to predict organoleptic factors such as flavour components are affected by the fact that a flavour may arise as a result of interactions among several constituents that are each present in minute amounts. All of these substances would be made up of the same combinations of O-H, C-H and N-H functional groups, which will again be dominated by the overwhelming presence of the same groupings in major constituents.

Yet another apparent anomaly is the ability to predict concentrations of heavy metals, such as lead, cadmium and mercury in some organic materials. Theoretically these types of prediction should be impossible because the metals have no absorbers in the NIR region. However, by their association with constituents such as cellulose, successful calibrations have been developed for their prediction in materials such as sediments (Malley & Williams, 1997).

The original assignments of absorbances caused by O-H stretching and other functional groups were identified by classical infrared spectroscopists, mainly in the 1930–1950 era. These workers gave small ranges over which the absorbances had been observed, and did not 'pin-point' a single nanometre (or reciprocal centimetre) at which an absorber would be expected to occur. Some of the absorbers were obtained by dissolving the substances, such as aliphatic hydrocarbons of different chain length, in carbon tetrachloride, which has no absorbers. A detailed list of absorbers is given in Table 2.3.

Table 2.3 Wavelength assignments

Range (nm)	Observed (nm)	Structure	Functional group	Material type
700–800	714	C-H (4th overtone)	C-H aromatic	Hydrocarbons aromatic
	747	C-H (4th overtone)	C-H methyl (CH_3)	Hydrocarbons aliphatic
	758			Cellulose, starch, oil
	762	C-H (4th overtone)	C-H methylene (CH_2)	Hydrocarbons aliphatic
	767	O-H (3rd overtone)	O-H as -CH_2-OH	Primary alcohols
	773	O-H (3rd overtone)	O-H as >CH-OH	Secondary alcohols
800–900	796	C-H combination	(C-H and δCH_3)	Hydrocarbons aliphatic
	803	C-H combination	(C-H and CH_3)	Hydrocarbons aromatic
	808			Protein
	813	C-H combination	(C-H and δCH_3)	Hydrocarbons aliphatic, oil
	816			Cellulose (dry)
	820			Starch (dry)
	830	C-H combination	(C-H and δCH_3)	Hydrocarbons aliphatic
	868			Protein
	876	C-H (3rd overtone)	C-H aromatic	Hydrocarbons aromatic
900–1000	908	C-H (3rd overtone)	C-H, CH_3	Hydrocarbons methyl
	930	C-H (3rd overtone)	C-H, >CH_2	Hydrocarbons methylene
	962	O-H	O-H no H-bonding	Alkyl alcohols
	964	O-H		O-H water Room temperature
	996	O-H	O-H, -CH_2-OH	Primary alcohols

Table 2.3 Wavelength assignments (continued)

Range (nm)	Observed (nm)	Structure	Functional group	Material type
1000–1100	1003	N-H	N-H	Aromatic amine
	1004			Cellulose (dry)
	1018			Protein
	1021	C-H combination	C-H (CH_3 and δCH_3)	Hydrocarbon aromatic & aliphatic
	1024	O-H	Possible H-bonding	Water 1–2°C
	1041	C-H combination	C-H (CH_2 and δCH_2)	Hydrocarbons aliphatic
	1047	O-H	O-H with H-bonding	Alkyl alcohols
1100–1200	1142-3	C-H	C-H aromatic	Hydrocarbons aromatic
	1156	O-H	O-H	Water in many materials
	1160	C = O	C = O carbonyl	Ketones aliphatic
	1170	C-H	HC = CH	Alkenes
	1195	C-H (2nd overtone)	C-H methylene	Hydrocarbons aliphatic
1200–1300	1211-5	C-H (2nd overtone)	C-H C-H, CH_2	Hydrocarbons aliphatic
	1225	C-H (2nd overtone)	C-H tertiary	Hydrocarbon aliphatic
1300–1400	1370-90	C-H combination	C-H CH_3 and δCH_3	Hydrocarbon aliphatic & aromatic
	1390	SiOH	SiOH	Silica
1400–1500	1410	C-H combination	C-H and δC-H	Linear aliphatic, e.g. oils
	1410	O-H (1st overtone)	O-H hydroxyl	O-H not hydrogen bonded
	1415	C-H combination	C-H methylene (CH_2)	Hydrocarbons, aliphatic
	1415	O-H	O-H alcoholic	Alcohols
	1420	O-H (1st overtone)	O-H aromatic	Phenolic O-H
	1430	N-H (1st overtone)	N-H primary amides	N-H primary amides
	1441	O-H (1st overtone)	O-H	Sugars, sucrose

Table 2.3 Wavelength assignments (continued)

Range (nm)	Observed (nm)	Structure	Functional group	Material type
1400-1500	1445	N-H (1st overtone)	N-H aromatic	Aromatic primary amine
	1450	C = O (1st overtone)	C = O > carbonyl	Ketones and aldehydes
	1450	O-H (1st overtone)	O-H -CH_2-OH	Starch, primary alcohols
	1460	N-H (1st overtone)	N-H	Urea
	1460	O-H (1st overtone)	O-H	Alkyl alcohols, water
	1463	N-H (1st overtone)	N-H -$CONH_2$	Amide/protein
	1483	N-H (1st overtone)	N-H -$CONH_2$	Amide/protein
	1490	O-H (1st overtone)	O-H	Cellulose
1500–1600	1500	N-H (1st overtone)	N-H	Amide/protein
	1500	O-H (1st overtone)	O-H	Alcohols, water
	1520	N-H (1st overtone)	N-H -$CONH_2$	Amide/protein
	1540	O-H (1st overtone)	O-H	Starch/polymeric alcohol
	1570	N-H (1st overtone)	N-H -CONH	Amide/protein
	1580	O-H combination	O-H	Alcohols, water
1600–1700	1620	C-H (1st overtone)	C-H =CH_2	Alkenes
	1680	C-H (1st overtone)	C-H aromatic	Classic reference wavelength
	1690-96	$CONH_2$	$CONH_2$	Proteins
	1694	C-H (1st overtone)	C-H_3	Hydrocarbons aliphatic
1700–1800	1705	C-H (1st overtone)	C-H CH_3	Hydrocarbons methyl
	1725	C-H (1st overtone)	C-H CH_2	Hydrocarbons methylene, oils
	1738	CONH	CONH	Proteins
	1740	S-H (1st overtone)	S-H thiols	Thiols
	1765	C-H (1st overtone)	C-H CH_2	Hydrocarbons, methylene
	1780	C-H (1st overtone)	C-H CH_2	Methylene, cellulose
	1790	O-H combination	O-H	Water

Table 2.3 Wavelength assignments (continued)

Range (nm)	Observed (nm)	Structure	Functional group	Material type
1800–1900	1820	O-H combination	O-H C-H comb	Cellulose
	1900	C-O (2^{nd} overtone)	C-O C = OOH	Carboxylic acids
1900–2000	1908	P-OH (1^{st} overtone)	O-H P-OH	Phosphate
	1920	C-O (2^{nd} overtone)	C-O (C = ONH)	Amide
	1930	O-H combination	O-H and HOH	Starch, cellulose
	1940	O-H combination	(O-H and H-O-H)	Water (classical filter reference)
	1950	C = O (2^{nd} overtone)	C = O (= OOR)	Acids and esters
	1960	O-H combination	(O-H and H-O-H)	Polysaccharides
	1980	N-H combination	N-H ($-CONH_2$)	Amides/protein
	1990	N-H combination	NH = $CONH_2$	Urea
	2000	N-H combination	N-H combination	Ammonia in water
2000–2100	2010	N-H combination	(R-C = O-NH_2)	Amides
	2030	N-H combination	NH_2-C = ONH_2	Urea
	2055	N-H combination	N-H/C = O as-CONH	Amidel: Amides, protein
	2060	N-H combination	N-H bend and N-H stretch combinaton amide II	Protein
	2070	N-H	N-H deformation	Urea
	2080	N-H combination	NH_2-C = O-NH_2	Urea
	2090	O-H combination	O-H polymeric	Complex carbohydrates
	2096	O-H	O-H deformation	Water or alcohols
	2100	C = O combination	C = O-O	Polysaccharides
	2100	C = O-O (3^{rd} overtone)	C = O and C-O stretch	Polysaccharides
	2100	O-H/C-O	O-H C-O combination	Polysaccharides

Table 2.3 Wavelength assignments (continued)

Range (nm)	Observed (nm)	Structure	Functional group	Material type
	2128	N-H/C = O	N-H/C = O combination	Polyamides
	2140	C-H/C = O	RC = CH and RC = O combination	Lipids
	2167	CONH	CONH peptide linkage	Proteins
	2170	C-H combination	-HC = CH	Alkenes
	2174	CONH	CONH combination	Proteins at low pH
	2180	N-H combination	N-H/C-N/C = O combination	Proteins
	2200	C-H C-O	C-H and C = O combination	Carbohydrates
	2230	CHO	CHO classic	Classic reference λ
	2270	OH combination	O-H stretch/ C-O stretch combinaton	Cellulose, glucose
	2280	C-H combination	C-H stretch + CH_2 deformation	Polysaccharides, starch
	2308	C-H combination	C-H methylene	Hydrocarbons aliphatic
	2310	C-H	C-H	Lipids
	2322	C-H combination	-C-H + CH_2	Polysaccharides, starch
	2332	C-H combination	-C-H + CH_2	Polysaccharides, cellulose
	2347	C-H combination	C-H + $R(CH_2)_nR$	Hydrocarbons aliphatic, lipids
	2352	C-H	C-H bending	Polysaccharides, cellulose
	2380	C-H combination	C-H stretch +C-C stretch combination	Lipids
	2407	C-H	C-H aromatic	Hydrocarbons aromatic

Table 2.3 Wavelength assignments (continued)

Range (nm)	Observed (nm)	Structure	Functional group	Material type
	2445	CONH	CONH peptides	Proteins
	2463	$CONH_2$	$CONH_2$	Peptides, proteins
	2470	C-N-C	C-N-C amide	Proteins
	2488	C-H combination	C-H + C-C combination	Cellulose
	2500	C-H combination	C-H/C-C/C-O-C combination	Polysaccharides

Source: Adapted from Workman and Weyer (2012)

References

AACC. 2000a. AACC International Approved Methods of Analysis. 11th Edition. Method 55-30.01. Particle Size Index for Wheat Hardness. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-55-30.01

AACC. 2000b. AACC International Approved Methods of Analysis. 11th Edition. Method 39-70.02. Near-infrared reflectance method for hardness determination in wheat. St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-39-70.02

Symes, K.J. 1965. The inheritance of grain hardness in wheat as measured by the particle size index. *Australian Journal of Agricultural Research*, 16, 113–123. https://doi.org/10.1071/AR9650113

Malley, D.F. & Williams, P.C. 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environmental Science and Technology*, 31, 3461-3467. https://doi.org/10.1021/es970214p

Suggestions for further reading

Dahm, D.J. & Dahm, K.D. 2007. *Interpreting Diffuse Reflectance and Transmittance. A theoretical introduction to absorption spectroscopy of scattering materials.* Chichester, UK: IM Publications.

Birth G.S. & Hecht H.G. 2001. The physics of near-infrared reflectance. In: P. Williams & K. Norris (eds). *Near-infrared Technology in the Agriculture and Food industries.* 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. pp. 1–15.

Colthup, N.B., Daly, L.H. & Wiberley, S.E. 1975. *Introduction to Infrared and Raman Spectroscopy.* 2nd Edition. New York, USA: Academic Press, Inc.

Workman Jr., J. & Weyer, L. 2012. *Practical Guide and Spectral Atlas for Interpretive Near-infrared Spectroscopy*. 2nd Edition. Boca Raton, Louisiana, USA: CRC Press, Taylor and Francis Group.

02

Reference Analysis

Just as the spectra form the heart of NIR technology, the reference test is the heart of NIRS application. Without the spectra, there is no technology at all. Without reference tests, there can be no quantitative applications, and even qualitative applications often require reliable reference data for the samples used in development of the training set. One major difference between reference and NIRS tests is that the NIRS method uses a much larger sample. Reference methods rarely use more than 1–2 g, and some use only a few micrograms or microlitres for the final assay.

This chapter is intended for users and potential users who aspire to obtain the best possible outcome from the application of their NIR instrument. It is particularly important to verify the accuracy and precision of reference testing before any NIRS work, because the NIRS results are always compared to those of the reference methods. Networked instruments are usually connected to a master instrument, and all adjustments to calibrations are made through the master instrument. Only the master instrument itself is calibrated to reference methods – there are legal as well as practical reasons for this.

3.1 Reliability of the reference method

Before attempting the development of a calibration, it is essential to determine the precision (standard error) of the reference testing that will be used to develop the calibration. One of the most difficult obstacles that instrument companies and their representatives have faced ever since the introduction of NIRS as a practicable method back in the early 1970s, has been the faith that companies and organisations have in their laboratories. When disagreements occur between reference and NIRS results, management and clients tend to believe the results of the 'tried and trusted' reference methods, rather than those of the 'black box' NIRS method.

Few users of NIRS pay enough attention to the reproducibility of their reference or NIRS testing. Poor results obtained by new or even seasoned NIRS users, are most frequently blamed on the instrument, whereas in many cases it is the reference method used for monitoring, or inefficient sample presentation that has caused the problem. The NIRS results are probably more reliable – once a calibration model has been developed and thoroughly evaluated, there are fewer areas for error in the NIRS than in reference 'wet chemistry' methods. This is because sample preparation is usually simpler, the sample size is much larger and the precision of spectral data on the repeat testing of any commodity is usually as good as, if not better than that of the reference test. The spectral data replace all of the chemical reagents, apparatus, addition of often small aliquots of sample and reagents, attention to conditions of reactions and other features of reference analysis, all of which are possible sources of error. The instruments are very stable, and the main sources of error in the NIRS test are the spectral precision, the integrity of the calibration, the natural heterogeneity of the sample and the operator through sample presentation.

This chapter will not recommend or describe reference methods in detail. It is intended to emphasise the need for accuracy and precision in the reference tests used in the development of calibration models (see Chapter 9) in order to identify areas of error in some tests, to encourage operators to study the precision of their reference test methods and to research areas where errors may be met. Error sources should be identified in any type of test. The chief concern is that the effectiveness of any reference method used in NIRS work influences the efficiency of the NIRS analysis. The tests described in more detail in this book are those most commonly used in grain and flour-mill laboratories. The principles of the identification of sources of error are, however, applicable to any test methods and any materials.

Many NIR instrument companies offer factory calibrations for a wide selection of applications. Clients should verify that these calibrations are satisfactory for their own uses. To do this, they should establish or identify a laboratory which they can rely on for reference data on samples that they submit for checking and monitoring their calibrations. The reproducibility of the selected laboratory can be determined by re-submitting samples under different identification from time to time. This is a necessary procedure. The laboratory will be using certified methods for reference testing, the subject of this chapter.

A reference test refers to any composition or functionality feature against which the spectral data of an NIR instrument is linked in the development of a calibration model. The first factors to which NIRS was applied historically were composition factors in grain – notably moisture, protein and oil contents. These are referred to as 'constituents' in this book. In terms of soil and manure analysis, moisture, total (elemental) nitrogen, phosphorus and carbon are among the most important constituents. For more than 40 years, NIR technology has also been applied to the prediction of functionality factors, such as kernel texture in wheat and digestibility in forages. Functionality factors are referred to as 'parameters' here, and refer to the physical and physicochemical attributes of the material to be tested that affect its end-use performance rather than chemical composition. Functionality of a material or commodity indicates if the material will actually do what it is supposed to.

The constituents most frequently tested for in agricultural materials, foods and feeds are: moisture, protein, starch and oil or lipids. These are considered 'easy' ones. In food and feed materials, fibre is also an important constituent. Even so, we come to the question of which fibre method should be used as reference test? There are five methods that are recognised for the determination of the fibre constituent. These are: crude fibre, neutral detergent fibre (NDF), acid detergent fibre (ADF), and total dietary fibre (TDF) that consists of NDF plus soluble dietary fibre (SDF).

Crude fibre is the constituent that is most widely-tested because food and feed companies are obliged to state the composition of their products for marketing purposes – crude fibre is legally recognised as the fibre content of the material. The other types of fibre are more directly related to its nutritional value, and are also tested by nutritionists in feed manufacturing companies. Total dietary fibre content tests are expensive and usually take 24 hours, so NIRS calibrations are particularly useful for this type of analytical work.

Functionality parameters such as metabolisable energy are further items for which successful calibrations have been developed. True metabolisable energy (TME) and apparent metabolisable energy (AME) are versions of metabolisable energy determined by different methods, and are applicable to different types of animals. The success of NIRS in predicting these more complicated factors has fired the ambition of the more innovative workers to attempt prediction of even less well-defined parameters, such as flavours in foods, and even the degree of freshness in fruits, vegetables and meat. For all of these parameters, the success of the application depends on the reliability of the reference method, the physical nature of the material and the sample presentation system.

3.2 Accuracy and precision

The reliability of a method includes accuracy and precision. Sometimes there is confusion about these two terms. 'Accuracy' implies 'closeness to the true value'. This is hard to achieve in agricultural commodities, because no constituents except water can be prepared in a pure state. Even water exists in different states in different materials (see the molecular concept in Chapter 2, section 2.3). For practical purposes, NIRS analysis accuracy means closeness to the reference value since that is the data which is used in the development of calibrations, as well as the values to which subsequent NIRS results will be equated.

Within a laboratory, accuracy can be established at the same time as precision through repeated analysis of a well-blended check sample. Provided that the reproducibility of the results is acceptable (see below), the mean result can be accepted as the accuracy for that laboratory. Accuracy can be assessed among all of the laboratories belonging to the same group by using the mean results of collaborative studies (ring tests) among these laboratories, using at least 6 samples with a range in composition. Provided that the reproducibility of all of the laboratories is acceptable, a consensus can be arranged among the participants that these mean results will be accepted as the accuracy that will form the basis of their future work.

'Precision' in NIRS testing means getting the same result every time. It includes repeatability and reproducibility. 'Reproducibility' includes all features of the test, including sub-sampling, sample preparation and presentation to the instrument, and testing by all of the operators likely to be doing the testing. It is determined by repeated analysis of the same sample (at least 10 times), including all of these steps, and all of the operators likely to be involved in future testing. The (standard deviation) SD of the results is the standard error of the test. 'Repeatability' includes all features of the test except sub-sampling and sample preparation.

It is determined by performing replicate tests (again, at least 10 repeats) on the same sample after sample preparation, and is a test of the method itself on the sample. The values are usually lower than for reproducibility, but should also be determined separately for all staff that are likely to use the methods – this will reveal the extent of individual operator error. Repeatability should be determined at the same time as reproducibility, because it is not possible to determine the error due to sub-sampling and sample preparation unless the error of the actual testing is known.

A useful exercise is to determine the relationship between the reproducibility and repeatability of a test method.

The percentage reproducibility attributable to repeatability can be calculated as follows:

$$[(\text{repeatability})^2 \times 100] \div (\text{reproducibility})^2$$

Thus if,

reproducibility = 0.231 and repeatability = 0.174

Then,

% reproducibility attributable to repeatability

$= [(0.174)^2 \times 100] \div (0.231)^2 = 56.7\%$

The sub-sampling and sample preparation steps contributed only 43.3% of the total error. This ratio is affected by the type of material. For example, the repeatability of testing wheat flour can be expected to be a lot lower than that of testing ground forage, or even whole grains.

One method of establishing the overall precision of the reference test is to select 12 to 20 samples and submit them for testing, some as single samples and others in duplicate. All aspects of the testing have to be involved, including sub-sampling where necessary, as well as sample preparation. At least four of the samples should be introduced as 'blind' duplicates. Table 3.1 gives an example of a suitable sequence, together with two sets of results, typical of excellent and of mediocre analytical performance. Five blind duplicates are included.

To calculate reproducibility, the first results of the respective 'blind' duplicates are used (e.g. 11.81 from the duplicates 11.81 and 11.84, 13.42 from the duplicates 13.42 and 13.57). This duplicate testing system allows one to determine the SD of duplicate testing on the same sample (repeatability; SD1) – the SD of the blind duplicate results also gives information on the overall reproducibility (SD2). In the case of the second example in Table 3.1 ('Mediocre result'), the SD1 (repeatability) values were similar for both series, showing that the error within

the laboratory after sample preparation was satisfactory. The higher value for SD2 (reproducibility) showed that sample preparation was an important source of error. The respective contributions of the repeatability to the overall error was 57.7% in the case of the 'Excellent' result, but only 12.8% for the 'Mediocre' overall error. This emphasises the importance of careful sample preparation on overall reproducibility. Information of this type is valuable in detecting areas where laboratory reference versus NIRS performance can be improved.

Table 3.1 Sequence for determination of laboratory performance for reference protein-testing of wheat

Laboratory ID	Sequence	Laboratory ID	Actual ID	Excellent Result	Mediocre Result
1	Single	1	**1**	14.26	14.31
2	Duplicate	2	**2**	11.81	11.72
3	Duplicate	2	**2**	11.77	11.82
4	Single	3	**3**	13.42	13.40
5	Duplicate	4	**4**	15.71	15.73
6	Duplicate	4	**4**	15.66	15.66
7	Single	5	**5**	10.43	10.51
8	Duplicate	6	**6**	12.93	12.87
9	Duplicate	6	**6**	13.10	12.79
10	Single	7	**2BD***	11.84	11.66
11	Duplicate	8	**7**	14.97	14.88
12	Duplicate	8	**7**	14.90	14.96
13	Single	9	**6BD**	**12.77**	13.00
14	Duplicate	10	**8**	9.50	9.56
15	Duplicate	10	**8**	9.56	9.64
16	Single	11	**9**	13.28	13.34
17	Duplicate	12	**3BD**	13.57	13.14
18	Duplicate	12	**3**	13.48	13.20
19	Single	13	**10**	14.38	14.16
20	Duplicate	14	**11**	12.16	12.48
21	Duplicate	14	**11**	12.24	12.33
22	Single	15	**12**	16.11	16.38
23	Duplicate	16	**10BD**	14.28	14.56

Table 3.1 Sequence for determination of laboratory performance for reference protein-testing of wheat (continued)

Laboratory ID	Sequence	Laboratory ID	Actual ID	Excellent Result	Mediocre Result
24	Duplicate	16	10	14.36	14.66
25	Single	17	4BD	15.74	15.41
26	Duplicate	18	13	12.55	12.41
27	Duplicate	18	13	12.48	12.31
SD1 (repeatability)**				0.061	0.067
SD2 (reproducibility)**				0.077	0.187

*'BD' = blind duplicate; the results of each 'blind' duplicate is shown in bold;
**SD difference = $\sqrt{\langle[\Sigma(x_n)^2]/(N \times 2)\rangle}$; Note: n = 1 to N; $\Sigma(x_n)$ = sum of all differences (a – b) from 1 to N (Mark, 1991)

The establishment of accuracy and reproducibility means a lot of repetitive and painstaking work, but it is the basis of any successful application of NIRS. Table 3.2 gives details of the calculation of the repeatability (SD1) and reproducibility (SD2) from Table 3.1 (Mark, 1991). It is emphasised that the accepted level of both the accuracy and reproducibility of any analytical method should be established at the beginning of any project. If the reproducibility is excellent, unless the accuracy has also been established, all future results could be reported precisely wrong.

Table 3.2 Method of calculation of the repeatability and reproducibility from the data of Table 3.1

Excellent results							
Repeatability (SD1)				Reproducibility (SD2)			
	Duplicate				Duplicate		
ID	**a**	**b**	**a – b**	**ID**	**a**	**b**	**a – b**
2, 3	11.81	11.77	0.04	**2, 10**	11.81	11.84	-0.03
5, 6	15.71	15.66	0.05	**4, 17**	13.42	13.57	-0.15
8, 9	12.93	13.10	-0.17	**5, 25**	15.71	15.74	-0.03
11, 12	14.97	14.90	0.07	**8, 13**	12.93	12.77	0.16
14, 15	9.50	9.56	-0.06	**19, 23**	14.38	14.28	0.10
17, 18	13.57	13.48	0.09		SD difference 2*		0.077
20, 21	12.16	12.24	-0.08		Mean difference 2		0.010
23, 24	14.28	14.36	-0.08				

Table 3.2 Method of calculation of the repeatability and reproducibility from the data of Table 3.1 (continued)

Excellent results							
Repeatability (SD1)				Reproducibility (SD2)			
	Duplicate				Duplicate		
26, 27	12.55	12.48	0.07				
	SD difference 1*		0.061				
	Mean difference 1		-0.008				
Mediocre results							
2, 3	11.72	11.82	-0.10	**2, 10**	11.72	11.66	0.06
5, 6	15.73	15.66	0.07	**4, 17**	13.40	13.14	0.26
8, 9	12.87	12.79	0.08	**5, 25**	15.73	15.41	0.32
11, 12	14.88	14.96	-0.08	**8, 13**	12.87	13.00	-0.13
14, 15	9.56	9.64	-0.08	**19, 23**	14.16	14.56	-0.40
17, 18	13.14	13.20	-0.06		SD difference 2*		0.187
20, 21	12.48	12.33	0.15		Mean difference 2		0.022
23, 24	14.56	14.66	-0.10				
26, 27	12.41	12.31	0.10				
	SD difference 1*		0.067				
	Mean difference 1		-0.002				

*SD difference = $\sqrt{\langle[\Sigma(x_n)^2]/(N \times 2)\rangle}$; Note: n = 1 to N;
$\Sigma(x_n)$ = sum of all differences (a – b) from 1 to N (Mark, 1991)

3.3 Approved reference methods

For contractual reasons, many commercial users of NIRS use as reference methods only those approved by associations such as the Association of Official Analytical Chemists (AOAC), the American Association of Cereal Chemists International (AACCI), the International Association of Cereal Science and Technology (ICC), the European Society of Brewing Chemists (ESBU), the American Society of Brewing Chemists (ASBC), the American Oil Chemists' Society (AOCS), the Australian Grain Science Association (AGSA) and others specific to different materials, such as forages, soils and pharmaceuticals. Analytical laboratories in other areas of industry will use certified methods that have been identified by the individual industries.

Although these methods are described in detail by the respective associations and societies, many of their users have implemented their own, usually minor, modifications to the methods, to suit their own equipment. It is important that

the users of any of these modified methods ensure that the results are exactly the same as the original methods, and that the changes to the methods as well as the results (before and after the changes) have been properly documented. Otherwise, contracts subject to cases of dispute can be declared invalid. This is particularly important if a laboratory seeks to gain certification under ISO 17025, the main standard used by testing laboratories.

During the past few years, several of the more frequently used Approved Methods of the AACCI have been the subject of re-examination by means of collaborative studies termed 'ring-tests'. The results of some of these are summarised in Table 3.3. All of the participants were well-established professional laboratories. The data of Table 3.3 are actual results. The reason for inclusion of this table is to show the levels of precision, as well as the errors that are encountered when a relatively simple test such as an ash test is carried out even by highly reputable laboratories, at least 8 of which were involved in all of the collaborative tests.

Table 3.3 Summary of collaborative tests of AACCI Approved Methods

Constituent	Method	Mean	Maximum	Minimum	SD	CV (%)
Moisture*	44-15.02	13.46	13.77	13.06	0.26	1.90
Ash	8-01.01	0.582	0.608	0.558	0.019	3.26
Ash**	8-02.01	0.611	0.678	0.561	0.043	7.04
Ash***	8-12.01	0.595	0.609	0.561	0.021	3.53
Starch damage	76-30.01	8.52	10.40	7.50	0.85	10.0
Protein	44-30.01	12.78	12.90	12.62	0.145	1.14
Wet gluten	ICC Standard 155	31.8	35.7	27.4	1.87	5.80
Gluten Index	ICC Standard 155	51.7	66.9	25.0	12.49	33.70

*Whole wheat, two-stage air-oven; **Rapid Magnesium Acetate method; ***Semolina method; CV = coefficient of variation

All of these tests are routinely carried out in flour mill and (large) bakery laboratories all over the world. The wet gluten and gluten index tests were carried out on the same samples. Ideally the coefficient of variation (CV) should be between 2.0 and 3.0% for approved analytical methods for composition. This was achieved in only two of these methods. The CV values of over 3% are questionable, while values of 7–10% are unacceptable. When this is observed, the reasons should be sought.

Table 3.7 gives guidelines for the CV that should be achieved for some commonly-used tests in cereal- and animal feed-mill laboratories. The CV-values for analysis of more complex materials, such as soils, may reach 10% or even higher.

Table 3.4 Coefficient of variation (CV) guidelines for some cereal test methods

Constituent	Method*	Acceptable CV (%)
Protein	46-16.01	1.5–2.0
Protein	46-30.01	1.0–1.5
Moisture: oven single-stage	44-15.02	1.5–2.0
Moisture: oven two-stage	44-15.02	2.0–3.0
Flour ash	8-01.01	2.5–3.5
Flour ash	8-02.01	3.5–5.0
Starch damage	76-31.01	2.5–3.5
Wet gluten	38-12.02	3.5–5.0
Crude fibre (feeds)	32-10.01	3.5–5.0

*All methods are taken from the AACCI method database

Some of the reasons for poor inter-laboratory reproducibility in AACCI Approved Methods are given in sections 3.3.1 to 3.3.3 for three of the cereal laboratory methods. It is important to research the possible causes for error in any test method used in an analytical laboratory. The examples explain errors that commonly occur in some methods carried out frequently in cereal laboratories. Many laboratories do not use the check sample method to check on their reproducibility. The cost of testing a check sample is the same as for testing any other samples, and the laboratory may consider check sample tests as a waste of time and money, and overlook the value of establishing and verifying their own reliability, or that of a laboratory that they choose to use as a reference laboratory.

3.3.1 Chief sources of error in Method 44-15.02 – Two-stage air-oven moisture

Undetected moisture loss during grinding at Stage A (if no pretreatment is carried out)

Method 44-15.02 (AACC, 2000a) recommends no pre-treatment (air-drying to lower moisture level) unless the moisture content is above 13% (as determined by moisture metre). Grinding samples in preparation for testing causes variable and undetermined loss in moisture – this can lead to undetected moisture loss of up to over 1% at 13% (see Chapter 8, Table 8.4).

Undetected moisture uptake during grinding in Stage B
(after pretreatment in Stage A)

If pre-treatment by air-drying brings the moisture level down below, e.g. 8–9%, the ground grain will be sufficiently hygroscopic to allow moisture to be absorbed from the atmosphere during sample preparation for Stage B.

General carelessness

The main reasons underlying carelessness are boredom and excessive speed. The repetitive routine of many analytical methods can induce boredom and lack of concentration. Supervisors need to remember that the members of their technical staff are all people. There are three parts to daily life – sleeping, living and working. If a task becomes sufficiently repetitive that the steps become automatic, there is a tendency for people to start to think about things of more interest in their lives. The drift in concentration can cause mistakes. Excessive speed is a more frequent cause of error. When a certain number of tests have to be completed in a normal day, factors such as late delivery of samples or unexpected telephone calls shorten the time available. The increased speed that is necessary to complete the day's workload can be an important cause of mistakes.

3.3.2 Chief sources of error in Methods 8-01.01 (wheat flour) and 8-02.01 (semolina) for ash

Ash is very hygroscopic. The few milligrams of ash resulting from a 5 g sample (only 25 mg at 0.5% ash) have a very large surface area and will rapidly absorb water even from within the desiccator. The ash is a better desiccant than $CaSO_4$ (Drierite). This is not a significant factor if only a few (up to 3) samples have been tested. However, with a larger number of samples, the repeated opening of the desiccator will change the relative humidity inside. An uptake of only 1 mg is a 4% error at an ash content of 0.50% – this will increase the apparent ash content from 0.50 to 0.52%. The rapid method 8-02.01 was developed to save time in flour mills (from 5–6 hours or overnight to 2 hours or less), and it is unusual to test more than one or two samples in duplicate so that moisture uptake in the desiccator is not a serious problem. Here non-uniform ashing is a more important source of error.

Method 8-01.01 – Basic method for flour ash content (AACC, 2000b)

1. Moisture uptake in desiccators
2. Number of samples
3. Moisture uptake during weighing of ash
4. General carelessness and excessive speed

Method 8-02.01 – Rapid magnesium acetate method for flour ash content (AACC, 2000c)

1. Incomplete and variable ashing due mainly to inefficient mixing of magnesium acetate with the flour
2. Moisture uptake in desiccators
3. Moisture uptake during weighing ash
4. General carelessness and excessive speed

3.3.3 Chief sources of error in Method 46-30.01 – Crude protein-combustion method for protein

Below are frequent sources of error in the Dumas (combustion) method for the determination of total nitrogen and protein contents in food and agricultural materials (Method 46-30.01 of the AACC International; AACC, 2000d).

1. Sampling and sample size
2. Sample preparation (grinding and especially blending after grinding)
3. Impure gases, especially oxygen (the main impurity is nitrogen) – Only the purest oxygen should be used (99.96%). This is the most expensive, but the best to use. The cost of the test must be considered in relation to the value of the result, which is invariably far greater than the few extra cents per test incurred by using the best oxygen available.
4. Inefficient pellet preparation (all nitrogen must be excluded) – The best way to prepare the pellet is to exclude any contaminating air (that contains about 70% nitrogen) by compressing and twisting the pellet by hand, possibly using a glove. The small pelleting devices that are supplied are less effective and may break the pellet during preparation.
5. Small amounts of moisture in reference chemical (e.g. EDTA)
6. Use of other than pure chemicals as reference – Again, the best chemicals should be used as reference. Pure EDTA has been found to contain small amounts of moisture. These are enough to change the calibration factor to the extent of giving significant day-to-day variability. Precision can be improved by drying the EDTA at 100°C before use. Tris-(hydoxymethyl)-amino-methane is a good reference, but is more difficult to dry consistently. Use of a material such as ground wheat as a reference is not recommended. This is a major and very variable source of error. Some laboratories use hippuric acid as a standard chemical, but it is not commercially available in such a high degree of purity as EDTA.
7. Moisture correction of data
8. Poor instrument maintenance
9. Sporadic use of instrument – Many instruments such as combustion analysers use capillary tubes and other small components in their construction. Experience has shown that they perform at their best by sustained use.

Performance suffers if the instruments are not used for even a few days. So, it is a good idea to analyse a few samples at least every other day, or to allow a day or two of consistent use between necessary down-times, before day-to-day analysis is resumed.

10. Insufficient reference or blank tests used during calibration
11. Instrument malfunction
12. General carelessness

Laboratories that contemplate use of NIR technology to replace some of their reference testing should carefully evaluate their reference methods to seek out and correct the sources of error. The early days of the use of NIRS in protein-testing of wheat in North America had the valuable side effect of focusing more attention on the accuracy and precision of Kjeldahl testing, and the importance of reporting protein results on constant moisture basis.

3.4 Arbitrary reference methods

Arbitrary methods are often developed to obtain information on materials such as fresh meat, fruit and vegetables. Analysis of these materials by standard methods is complicated by their high moisture content and variable physical size, shape and general characteristics. A basic issue is that of moisture content. Most materials contain moisture and to avoid errors and misinterpretation, it is important to report results on a constant moisture basis (usually moisture-free). In the application of NIRS to materials with varying moisture content, it is necessary to determine the composition on a constant moisture basis. For direct application of NIRS, the analytical results must be re-computed to the original moisture content of the material before use in calibration development. Laboratory reports of analytical moisture tests results should include both as-is and constant moisture content data.

Many laboratories develop their own methods of determining functionality parameters to give them the information they need. In the case of such methods, absolute accuracy cannot be determined. The method simply has to provide the information sought. Though an experienced technician can get results of excellent reproducibility with the method, it does not guarantee that it is suitable for long-term application, or for use as a reference method for NIRS model development. At the very least, experienced technicians will need vacations, experience sickness and eventually retire. It is imperative that such methods are made as free as possible of error sources, and that other technicians are equally competent with the method. If an arbitrary method proves to be reliable over time it could

be submitted for official approval by a recognised body, such as the AOAC. Many of such methods are time-consuming, and their users quickly realise the potential of NIRS to save turn-around time and cost per test. Sample presentation is a key factor in these NIRS applications.

3.5 Last words on reproducibility

Reproducibility of results is the most important criterion. The reproducibility of arbitrary methods must be established in the same way as any other methods intended for use as reference methods for NIRS model development. This can be done by repeated analysis of a check sample or samples, and may not reach the level of precision as the approved methods. Identification of sources of variance and error in such tests, as well as sample assembly also follow the same rules as any other method. Obtaining and maintaining check samples of fresh materials, such as meat, fish, milk and other perishable materials is recognised as being difficult, but a reasonable substitute is to retain one sample from early testing each day, protect it as well as practicable and re-test it a few times during the day in order to establish the degree of reproducibility. Over a period of a few days an average value for the reproducibility can be obtained. The reproducibility of NIRS testing is usually superior to that of reference testing. This is particularly true in the case of continuous on-line testing by NIRS, which eliminates the need for sampling and associated errors.

References

AACC. 2000a. AACC International Approved Methods of Analysis. 11th Edition. Method 44-15.02. Moisture – air-oven methods. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-44-15.02

AACC. 2000b. AACC International Approved Methods of Analysis. 11th Edition. Method 08-01.01. Ash – basic method. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACC IntMethod-08-01.01

AACC. 2000c. AACC International Approved Methods of Analysis. 11th Edition. Method 08-01.01. Ash – rapid (magnesium acetate) method. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-08-02.01

AACC. 2000d. AACC International Approved Methods of Analysis. 11th Edition. Method 46-30.01. Crude protein – combustion method. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-46-30.01

Mark, H. 1991. *Principles and Practice of Spectroscopic Calibration.* New York, USA: John Wiley. p. 144.

4

Statistical terms for evaluation of accuracy and precision

Most people are put off by the word 'statistics'. Two things are important here: how to apply the formulae and how to interpret the results. There is no need to know the complex mathematics or how these formulae were derived – this expertise has been incorporated into software used in the calibration and monitoring of NIR instruments.

Before discussing the statistical terms, there are *five important messages* to all NIRS users.

First, the values of X and Y used in this book will usually refer to reference and NIRS data, but can apply to any two sets of data. The **second** is that a single result for any test on any sample is called an 'observation', and the total number of samples tested, or data evaluated is called the 'population' in this book. There are two types of population. The first type is the population of whatever is there, e.g. all of the bales of hay in a field. Sample selection is not a factor because all of the samples are included and have to be tested. The second type is the population as defined by the samples that have been assembled to test a hypothesis, for example: What is the correlation between maize hardness and milling quality? The success of this type of exercise depends on the selection of samples. Practical

application will rely heavily on the results of the statistical analysis, and sample selection and assembly are also crucial. Unless the sample selection and assembly have been carefully considered, and the samples truly represent the variance likely to be met, biases and outliers could occur. Decisions and actions based on the statistics may thus not be valid.

The **third** message is to look at all of the data. The statistics are all computed from them. Ideally, this should be done as the data are being generated, which takes place during daily analysis or when recording spectra for calibration model development (using a computerised spectrophotometer). If you are doing day-to-day analysis you should keep an eye on the results to make sure that they look 'familiar' and that there are no sudden major changes. For example, if the results for a certain test normally lie between 11 and 15% and a result of 9 or 17% appears, it should be checked. If a value of, e.g. 23.7% appears in the same series of testing, it is more likely to be a clerical error than an analytical one.

The **fourth** message is that in this section there are a lot of formulae. These are included in 'long-form' detail so that the reader can follow the reasoning underlying the calculations. Different versions of these formulae occur, which are more familiar to chemometricians, but the formulae presented in this book give all of the factors needed to calculate the statistics. In practice nowadays, few people actually calculate statistics such as standard deviations, or coefficients of correlation or regression since statistical packages include all the formulae – users only have to enter their data.

Finally, the **fifth** message. This draws attention to the difference between accurate analysis and screening. Accurate analysis by NIRS requires an r^2-value of 0.95 or better. But values of even as low as 0.70 can provide an adequate way of screening large numbers of samples into high, low and intermediate levels of acceptability in industry. Guidelines can be established for upper and lower limits of acceptability. All samples above the high limit can be safely accepted, those below the lower limit, rejected or down-graded, while the samples that fall between these limits will need testing by established methods. The NIRS screening can affect up to 67% reduction in the expense and time of application of conventional test methods to large populations of samples, such as the arrival of loads of grain at a grain terminal by train.

4.1 Useful statistical terms for evaluating NIRS calibrations

Experienced operators can tell whether their instrument is performing properly just by 'eyeballing' the results, but for reporting purposes the efficiency of an NIRS

calibration is usually evaluated and reported on the basis of applied statistics. Several terms are needed for the interpretation of statistical analysis of the results of NIRS testing. Unless all of them are understood and correctly appraised, the operator may draw the wrong conclusions leading to frustrating and sometimes costly discrepancies. Important terms include the following:

1. mean;
2. standard deviation (SD) – an estimate of the degree of homogeneity, or heterogeneity of the population;
3. coefficient of variation (CV) – which relates the SD to the mean;
4. bias – the mean difference between any two sets of data;
5. correlation coefficient (r) – indicates an association between the X (independent variable) and Y (dependent variable) data;
6. coefficient of determination (r^2) – a measure of how close the data are to the fitted regression line, or to assess how well a model explains and predicts future outcomes;
7. regression coefficient (b) and intercept (a) from the simple linear regression equation;
8. distribution of differences between x and y data – in NIRS work, these are usually NIRS predicted and reference results;
9. standard error of a single test (SET) or analytical precision – also referred to as the standard error of the laboratory (SEL);
10. standard error of prediction (SEP) – SD of differences between NIRS predicted and reference values in the validation sample set;
    standard error of calibration (SEC) – SD of differences between NIRS and reference values in the calibration sample set;
11. standard error of cross-validation (SECV);
12. root mean square error of prediction (RMSEP) – also termed root mean square deviation of differences (RMSD); and
13. ratio of the SEP to the SD_y (RPD) – the Ratio of (standard error of) Prediction to (standard) Deviation or RPD.

4.1.1 Mean

There are three forms of the mean. The most common is the **arithmetic mean**. It is calculated by summing all values of x (or y) and dividing by N, the number of observations. It is symbolised by x-bar ($\bar{x}$) or y-bar ($\bar{y}$). The formula for computing the mean is:

$$\text{Mean } (\bar{x}) = \Sigma x_n / N$$

Where, Σx_n = the sum of all x-values and n = 1 to N

Table 4.1 gives an example of calculating the arithmetic mean. The data are typical results for wheat protein content in percentage (%), but the principle applies to any population of data.

Table 4.1 Example of calculation of the mean

Observation	Data (x) Wheat protein content (%)	Statistic	Value
1	9.7	N	12
2	11.9	Σx_n	157.0
3	17.3	Mean	13.08
4	14.2		
5	12.6		
6	10.3		
7	15.1		
8	14.8		
9	11.1		
10	12.4		
11	13.5		
12	14.1		

Note: Σx_n = the sum of all x-values from 1 to N

The second form of the mean is the 'running mean'. The running mean is based on a sequence of e.g. 12 observations. When the next (13^{th}) observation is made the first observation is left off, and the running mean is calculated from the 12 that have been retained. Table 4.2 gives an example of calculation of the running mean. It is used to maintain a record of, e.g. the protein content of increments of wheat during loading of a ship (assuming that the increments are of approximately equal weight). The sequence here is 12 samples. When the 13^{th} sample is added the original first value (12.8) is omitted, and the original 2^{nd} sample becomes the first in the second sequence, and so on.

Table 4.2 Example of calculation of the running mean

N (1^{st} sequence)	Protein (%)	N (2^{nd} sequence)	Protein (%)
1	12.8	2	13.6
2	13.6	3	13.1
3	13.1	4	13.5

Table 4.2 Example of calculation of the running mean (continued)

N (1st sequence)	Protein (%)	N (2nd sequence)	Protein (%)
4	13.5	5	13.2
5	13.2	6	13.7
6	13.7	7	12.9
7	12.9	8	14.4
8	14.4	9	13.3
9	13.3	10	13.4
10	13.4	11	13.5
11	13.5	12	13.3
12	13.3	13	13.8
Sum	160.7	Sum	161.7
Mean	13.39 (13.4)	Running Mean	13.48 (13.5)

The third form of the mean is the 'weighted mean'. The weighted mean is used to calculate the average of a series of observations, such as protein content based on, for example, the different quantities of grain of different protein contents that may be added to a silo or cargo. It is calculated by multiplying the weight of each increment (such as farmers' deliveries to a country elevator) by the protein content, then dividing the total (weight × protein) by the total (weight). Table 4.3 gives an example of calculation of the weighted mean.

Table 4.3 Example of calculation of the weighted mean

N	Protein (%)	Weight (tonnes)	Weight × Protein	Cumulative Protein
1	12.8	14.6	186.88	12.80
2	13.6	21.2	288.32	13.27
3	13.1	20.3	265.93	13.21
4	13.5	44.6	602.10	13.33
5	13.2	28.7	378.84	13.31
6	13.7	16.9	231.53	13.35
7	12.9	33.3	429.57	13.27
8	14.4	24.9	358.56	13.41
9	13.3	18.3	243.39	13.40

Table 4.3 Example of calculation of the weighted mean (continued)

N	Protein (%)	Weight (tonnes)	Weight × Protein	Cumulative Protein
10	13.4	26.6	356.44	13.40
11	13.5	37.5	506.25	13.41
12	13.3	29.6	393.68	13.40
Sum	-	316.5	4241.49	

Weighted mean protein content = $\sum$ (weight$_n$ × protein$_n$) / total weight

4.1.2 Standard deviation (SD)

The SD is an expression of the variability, or variance in the data. It is the square root of the variance. One formula for computing the SD_x (or SD_y) is:

$$SD_x = \sqrt{\langle\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\}/(N-1)\rangle}$$

Where, n = 1 to N

Σx_n = sum of all x-values from 1 to N.

Table 4.4 Details of two alternative methods for calculating the standard deviation (SD)

Observation	Data (x)	Data-squared (x^2)	Alternative method	Difference (x - mean)	Difference-squared
1	9.7	94.09		-3.383	11.445
2	11.9	141.61		-1.183	1.400
3	17.3	299.29		4.217	17.783
4	14.2	201.64		1.117	1.248
5	12.6	158.76		-0.483	0.233
6	10.3	106.09		-2.783	7.745
7	15.1	228.01		2.017	4.068
8	14.8	219.04		1.717	2.948
9	11.1	123.21		-1.983	3.932
10	12.4	153.76		-0.683	0.466
11	13.5	182.25		0.417	0.174
12	14.1	198.81		1.017	1.034
N		12			
$\bar{x}$ (mean)		13.08			
Σx_n		157.00	$\Sigma(x_n - \bar{x})$	0.04	
Σx_n^2		2106.56	$\Sigma(x_n - \bar{x})^2$		52.4768

Table 4.4 Details of two alternative methods for calculating the standard deviation (SD) (continued)

Observation	Data (x)	Data-squared (x^2)	Alternative method	Difference (x – mean)	Difference-squared
$(\Sigma x_n)^2/N$		2054.0833	$[\Sigma(x_n - \bar{x})^2]/(N-1)$		4.7706
$\Sigma x_n^2 - [(\Sigma x_n)^2/N]$		52.4767	$\sqrt{\langle[\Sigma(x_n - \bar{x})^2]/(N-1)\rangle} = SD$		2.18
$\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\}/(N-1)$		4.7706			
$\sqrt{\langle\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\}/(N-1)\rangle} = SD$		2.18			

Note: n = 1 to N; Σx_n = the sum of all x-values from 1 to N

Table 4.4 gives an example of calculation of the SD by two methods. The SD is also the square root of the sum of the squares of the differences between each value and the mean (because some of the differences will be positive and the others negative it is necessary to square the differences), divided by N–1. Figure 4.2 is a graphical representation of the SD.

An important point, with regard to Figure 4.1, is that only about 68% of the total population falls between one SD above and below the mean. The values of the mean ±1.96 times the SD are referred to as the 95% confidence limits. For practical purposes the 95% confidence limit is a better guideline to the variance of a population than is the SD. Figure 4.1 illustrates this for a normal (Gaussian) distribution of data.

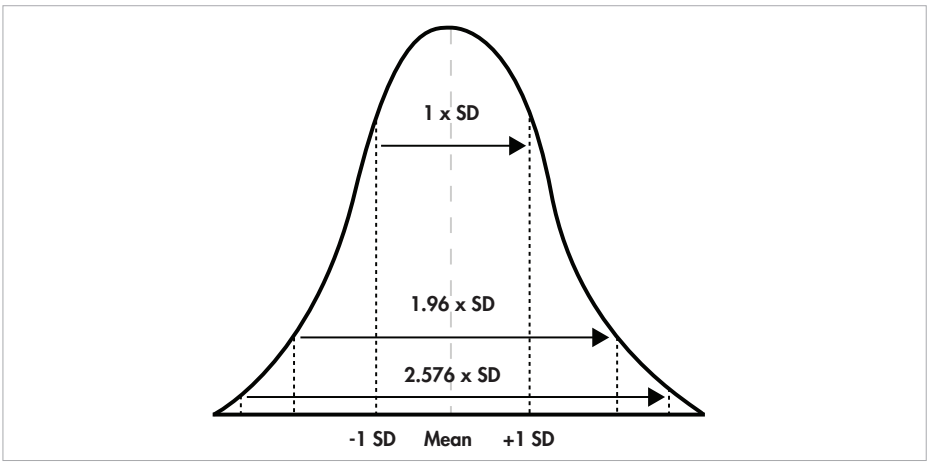


Figure 4.1 Illustrating the standard deviation (SD). The mean ±1.96 times the SD are referred to as the 95% confidence limits.

But why N–1? N–1 is called the 'degrees of freedom'. This is the number of observations in a population that are able to vary. Consider a set of 12 samples

(N = 12), with a mean of 13.5% protein content. No matter how variable the individual data of the first 11 samples, for the mean to be 13.5 the 12th value must be 13.5, and only 11 of the observations are free to vary. Table 4.5 gives examples of the true meaning of the SD for protein content and Falling Number (FN), two commonly tested parameters in wheat. To account for 99.9% of the variance in the population the SD is multiplied by 3.0.

Using the 99% confidence limits, the range between which 99% of the samples will fall is surprisingly large, even at the low SDs (standard errors per test) of 0.15% for protein and 20 seconds for the FN. Remember also that 1% of the population will fall outside of these limits. This means that in a population of 200 and a SD of 0.15, 2 samples will fall outside of a range of 12.9 ± 0.39 %. These are not outliers – they are part of the population.

Table 4.5 Implication of the standard deviation (SD)

Parameter	Mean ± SD		Mean ± 1.96*SD		Mean ± 2.576*SD		
Percent of population (%)	68		95		99		
Protein content (%)	Low	High	Low	High	Low	High	Range
Mean = 12.9; SD = 0.15	12.75	13.05	12.61	13.19	12.51	13.29	0.78
Mean = 12.9; SD = 0.25	12.65	13.15	12.41	13.39	12.26	13.54	1.28
Falling Number(s)							
Mean = 350; SD = 20	330	370	311	389	298	402	104
Mean = 350; SD = 25	325	375	301	399	286	414	128

4.1.3 Coefficient of variation (CV)

The CV is the population SD x 100 divided by the population mean ($\bar{x}$):

$$CV = (SD*100)/Mean\ (\bar{x})$$

It is expressed as a percentage. It relates the SD to the mean and provides a relative measure of variability in the population that is independent of the unit of measurement used. Table 4.6 gives an example of calculation of the CV based on data from Table 4.4.

Table 4.6 Example of calculation of the coefficient of variation (CV)

Observation	Data	Statistic	Value
1	9.7	N	12
2	11.9	Mean ($\bar{x}$)	13.08
3	17.3	Σx_n	157.0
4	14.2	Σx_n^2	2106.56
5	12.6	SD	±2.18
6	10.3	CV (%)	16.67
7	15.1		
8	14.8		
9	11.1		
10	12.4		
11	13.5		
12	14.1		

Note: Σx_n = sum of all x-values from 1 to N

Table 4.7 Typical results for repeatability and reproducibility for testing grain for protein content (%)

Test	Repeatability	Reproducibility
1	13.28	13.18
2	13.35	13.45
3	13.64	13.84
4	13.58	13.38
5	13.42	13.42
6	13.61	13.71
7	13.47	13.37
8	13.33	13.13
9	13.67	13.77
10	13.29	13.39
11	13.59	13.64
12	13.38	13.33
Mean (%)	13.47	13.47
SD	0.143	0.216
CV (%)	1.07	1.60

The above sections on Mean, SD and CV (%) are illustrated by examples of data for incoming deliveries of wheat (or similar grains or crops), where the variance can be high. Table 4.7 gives data more typical of the results for the repeatability and reproducibility (see Chapter 3, section 3.2) of testing a check sample for protein content. Based on reproducibility data, the 95% confidence limits for protein content would be ±0.432 (0.43), and the overall range in results that could be expected would be 1.30% (±0.216 x 3 = 0.65). The extreme results of 13.13 and 13.84 are not outliers. The data in Table 4.7 are typical of NIRS testing for protein content in whole wheat kernels. Determination of repeatability in whole-grain NIR instruments for which the sample is introduced using a hopper is not practicable because of the method of sample presentation.

When the CV is used to examine variance among replicate analyses on the same sample, it can be used to express and report the reproducibility of the method. The size and interpretation of the CV depends partly on the source of the data. For quality assurance applications, the CV of reference testing for constituents such as protein and moisture contents should be about 1.0–1.5%. Values of the CV between 2 and 3% are acceptable for some tests (see Chapter 3, Table 3.7). Table 4.8 gives some guidelines for the interpretation of the CV. For some applications such as analysis of soils and manures, that are very variable, calibrations have been used with CV values as high as 8–10%.

Table 4.8 Interpretation of the coefficient of variation (CV) for replicate determinations on a sample in three situations

CV-value	Interpretation of reference test results (e.g. protein content)	Interpretation of NIRS-predicted constituents	Interpretation of NIRS-predicted functionality
0.5–1.0	Exceptional	*	*
1.1–2.0	Excellent	Exceptional	*
2.1–3.0	Very good	Excellent	Exceptional
3.1–4.0	Good	Very good	Excellent
4.1–5.0	Fair	Good	Very good
5.1+	Needs investigation	Fair	Good

*Unlikely to be achieved

The mean and CV are accessories to the SD and SEP (see sections 4.1.2 and 4.1.10). The CV will be inflated when the reference value for the check sample is low, and when the SEP is constant over the concentration range. In cases where check samples extend over a large range of concentration, it is often better to report the SEP than the CV. This is illustrated in Table 4.9 for two constituents, protein

and deoxynivalenol (DON) contents in wheat. The CV of 2.5% is considered to be high for protein predictions in whole wheat or flour. The CV of 7.5% for prediction of DON is also high, but not uncommon for prediction of parameters such as DON.

Table 4.9 Influence of concentration on the CV (coefficient of variation) at constant SEP (standard error of prediction)

Constituent / SEP	Reference result (mean values)					
Protein (%) / SEP = 0.20	8.0	10.0	12.0	14.0	16.0	18.0
CV (%)	2.50	2.00	1.67	1.43	1.25	1.11
DON ppm / SEP = 0.60	0.25	0.50	1.00	2.00	4.00	8.00
CV (%)	240	120	60	30	15	7.5

The SEP should be determined at high and low levels of composition because the efficiency of prediction may be higher or lower at different levels. The same applies to determination of the SET method (see section 4.1.9), because the efficiency of reference methods may also change at different levels of concentration. The CV is not really applicable to calculations such as the SD of differences between duplicates. The mean difference is likely to be low, which would inflate the CV unrealistically, or even 0.00, which would make the CV impossible to compute.

4.1.4 Bias

The bias in a series of comparisons is the difference between the independent and the dependent variable. When calculated from the NIRS predictions of data in the validation sample set, the bias is the mean of the differences between the NIRS and reference data, and is an important part of the overall accuracy of the calibration (the SEP is the other component of accuracy). In the world of commerce and industry, the bias is one of the most important statistics. When payments are being made on the basis of composition, such as premiums or discounts for protein or moisture content, biases mean money. A bias of as little as 0.1% can incur considerable profits or losses when large volumes of a commodity are involved. Also, when NIRS data are being used to formulate feed mixes and so on, biases can cause inaccuracies in the composition of the final products, which can change the nutritional quality, and subsequent changes in animal productivity. Biases can occur even when the correlation coefficient and SEP statistics indicate that an excellent calibration has been developed. The main causes of bias include changes in the source of raw materials, grains or other commodities from different locations or seasons, changes in processing conditions, changes in ambient temperature and humidity, as well as other factors.

The formula for calculating bias (which can be positive or negative) is:

$$\text{Bias} = (\Sigma x_n / N) - (\Sigma y_n / N)$$

Where, x_n and y_n can be any two sets of data

$n = 1$ to N

Σx_n = sum of all x-values from 1 to N

Σy_n = sum of all y-values from 1 to N

4.1.5 Correlation coefficient (r)

The r shows the degree to which two sets of data, x_n and y_n (n = 1 to N), e.g. the NIRS and reference results, agree with each other. Perfect agreement, with no difference at all between the two data sets will result in an r of 1.000. In practice this is impossible since a certain amount of error in both X and Y data are unavoidable. The x_n and y_n data may be either positively or negatively correlated. Figure 4.2 illustrates how the correlation plot will change, depending on the value of r. Table 4.10 gives guidelines for the interpretation of r.

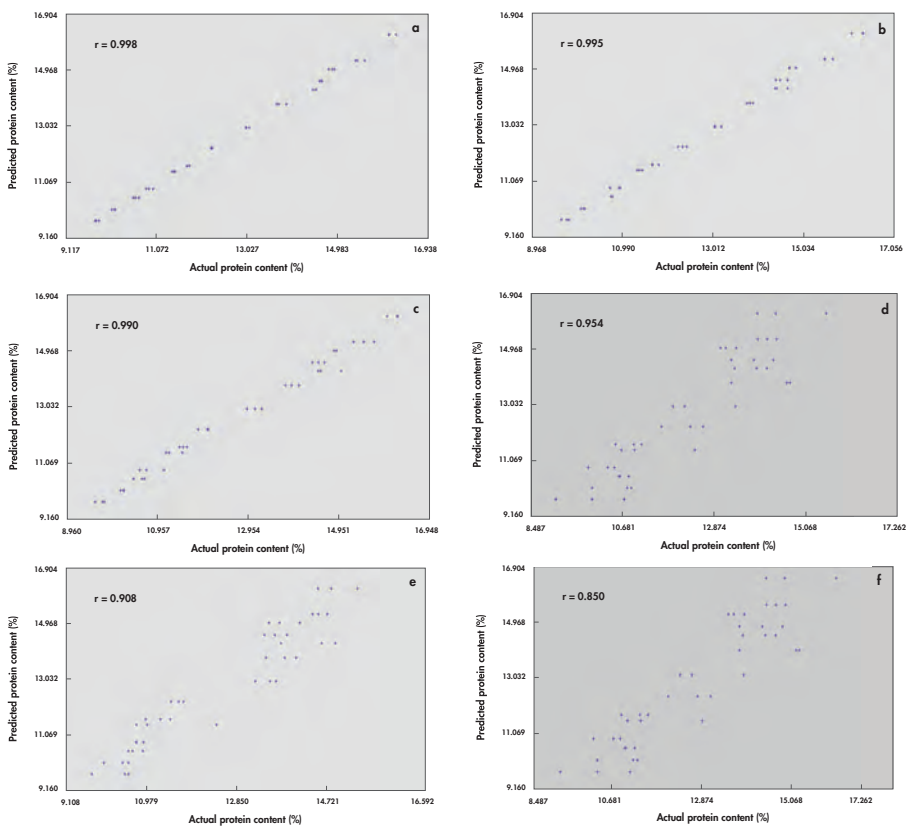


Figure 4.2 Illustration of how the correlation plot will change, depending on the value of the correlation coefficient (r).

The formula for computing r is:

$$r = \frac{\Sigma(x_n \times y_n) - [(\Sigma x_n \times \Sigma y_n)/N]}{\sqrt{\langle\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\} \times \{\Sigma y_n^2 - [(\Sigma y_n)^2/N]\}\rangle}}$$

Where, $n = 1$ to N

Σx_n = sum of all x-values from 1 to N

Σy_n = sum of all y-values from 1 to N

4.1.6 Coefficient of determination (r^2)

It is important to distinguish between the use of R or r. Authors may use either R and R^2, or r and r^2. R is normally associated with multiple linear regression or correlation, and r with simple linear regression or correlation. When reporting the results of correlations between NIR and reference data, e.g. in the evaluation of a calibration model, this is a simple linear relationship, and r or r^2 is the correct form to use.

The coefficient of determination is given by r^2. It shows the proportion of the variance of the NIRS predicted data that can be predicted by the spectral or log 1/R data, and shows the degree to which the predicted data can be expected to change, for a given change in the spectral data. For example, an r-value of 0.97 gives r^2 of 0.941. This means that 94.1% of the variance in y_n ($n = 1$ to N) can be predicted by x_n ($n = 1$ to N). It follows that only a combined 5.9% of the variance in x_n is attributable to other factors, such as sample preparation, reference testing and so on. Values of r^2 are always positive, regardless of the sign of r. For screening purposes, an r-value of 0.71 can be considered as a cut-off point. The r^2 value is 0.504. This means that one single variable explains slightly more than 50% of the total variance, so that all other factors combined cannot exceed the importance of that one variable. For practical purposes r^2 gives more information, and is preferred to r. Guidelines for the interpretation of both r and r^2 are given in Table 4.10.

A word of caution about r. Some authors report the statistical significance of r. The statistical significance of r can be calculated by the formula:

$$t = \frac{r}{\sqrt{\langle(1 - r^2)/(N - 2)\rangle}}$$

The value of t (obtained from statistical tables) depends on the number of degrees of freedom, which depends on the number of observations. If the number of observations is very large, the term $(1 - r^2)$ is divided by a large value, and

the value of t can appear to be statistically significant at quite low values of r. Table 4.11 shows how the significance of r at the 95% significance level (P = 0.05) changes as the number of observations increases:

Note that with a population of 60 samples or more, the value of an r of only 0.26 (r^2 = 0.068) is statistically significant at the P = 0.05 level. However, an r^2-value of 0.068 leaves over 93% of the residual variance unexplained. So, while the correlation coefficient may be statistically significant, it has little practical value in relating the x to the y variables.

Table 4.10 Guidelines for the interpretation of r and r^2

r	r^2	Interpretation
Up to 0.5	**Up to 0.25**	Not usable in NIRS calibration
0.51–0.70	**0.26–0.49**	Poor correlation, reasons should be researched
0.71–0.80	**0.50–0.64**	Acceptable for very rough to rough screening
0.81–0.90	**0.66–0.81**	Acceptable for screening and some other 'approximate' calibrations
0.91–0.95	**0.83–0.90**	Usable with caution for most applications, including research
0.96–0.98	**0.92–0.96**	Usable in most applications, including quality assurance
0.99+	**0.98+**	Excellent, usable in any application

Table 4.11 Levels of significance of r at different values of N

N	t (P = 0.05)	r (P = 0.05)	r^2	% residual variance*
12	2.18	0.58	0.336	66.4
15	2.13	0.52	0.271	72.9
20	2.09	0.45	0.202	79.8
24	2.06	0.41	0.168	83.2
40	2.02	0.32	0.102	89.8
60	2.00	0.26	0.068	93.2
100	1.99	0.20	0.040	96.0

*Percentage of residual variance not accounted for

4.1.7 Regression coefficient (b) and intercept (a)

The regression coefficient (b), also called the 'slope', is a useful indication of the potential effectiveness of the calibration. The slope shows the rate of change of

Y as a function of the rate of change in the other variable (X). The slope, together with the intercept (a) shows the degree to which values of Y can be predicted from those of X, or vice versa. In a perfect relationship between X and Y (where both X and Y refer to the same types of data, e.g. protein content by NIRS and reference methods), the values of 'r' and 'b' will both be 1.000, and 'a' will be 0.000. Again, a certain amount of error is unavoidable, so that the slope will be more or less than 1.000, and the intercept will differ from zero. In NIR technology, the error level of the spectral data is usually lower than that of the reference data. As in the case of the correlation coefficient, b and a can also be positive or negative. Table 4.12 provides full details for the calculation of r, b and a.

The formula for computing b_{yx} is:

$$b_{yx} = \frac{\Sigma(x_n \times y_n) - [(\Sigma x_n \times \Sigma y_n)/N]}{\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\}}$$

Where, n = 1 to N

Σx_n = sum of all x-values from 1 to N

Σy_n = sum of all y-values from 1 to N

The formula for computing a is:

$$a = \bar{y} - (b_{yx} \times \bar{x})$$

Table 4.12 Details for calculating coefficient of correlation (r), regression coefficient or slope (b_{yx}) and intercept (a_{yx})

Observation	x-data	x-data-squared (x^2)	y-data	y-data-squared (y^2)	x-data × y-data ($x \times y$)
1	9.7	94.09	10.0	100.00	97.00
2	11.9	141.61	11.8	139.24	140.42
3	17.3	299.29	17.1	292.41	295.83
4	14.2	201.64	14.6	213.16	207.32
5	12.6	158.76	13.0	169.00	163.80
6	10.3	106.09	10.0	100.00	103.00
7	15.1	228.01	14.8	219.04	223.48
8	14.8	219.04	15.1	228.01	223.48
9	11.1	123.21	11.7	136.89	129.87
10	12.4	153.76	12.9	166.41	159.96

Table 4.12 Details for calculating coefficient of correlation (r), regression coefficient or slope (b_{yx}) and intercept (a_{yx}) (continued)

Observation	x-data	x-data-squared (x^2)	y-data	y-data-squared (y^2)	x-data × y-data ($x \times y$)
11	13.5	182.25	13.1	171.61	176.85
12	14.1	198.81	13.8	190.44	194.58
N = 12	$\Sigma x_n = 157.0$	$\Sigma x_n^2 = 2106.56$	$\Sigma y_n = 157.9$	$\Sigma y_n^2 = 2126.21$	$\Sigma(x_n \times y_n) = 2115.59$
	$\bar{x} = 13.08$		$\bar{y} = 13.16$		
	$SD_x = 2.18$		$SD_y = 2.10$		
	$(\Sigma X)^2/N = 2054.08$		$(\Sigma Y)^2/N = 2077.70$		$(\Sigma X \times \Sigma Y)/N = 2065.86$
	$\Sigma X^2 - (\Sigma X^2/N) = 52.48$		$\Sigma Y^2 - (\Sigma Y^2/N) = 48.51$		$\Sigma(X \times Y) - [(\Sigma X \times \Sigma Y)/N] = 49.73$
$\sqrt{\langle[\Sigma x_n^2 - (\Sigma x_n^2/N)] \times [\Sigma y_n^2 - (\Sigma y_n^2/N)]\rangle} = \sqrt{\langle 52.48 \times 48.51\rangle} = 50.45$ $r = \{\Sigma(x_n \times y_n) - [(\Sigma x_n \times \Sigma y_n)/N]\}/\sqrt{\langle[\Sigma x_n^2 - (\Sigma x_n^2/N)] \times [\Sigma y_n^2 - (\Sigma y_n^2/N)]\rangle} = 49.73 / \sqrt{50.74} = 0.9857$					
$b_{yx} = \{\Sigma(x_n \times y_n) - [(\Sigma x_n \times \Sigma y_n)/N]\}/\{\Sigma x_n^2 - [(\Sigma x_n)^2/N]\} = 2115.59/2065.86 = 0.9477$					
$a_{xy} = (\bar{y}) - (b_{yx} \times \bar{x}) = (13.16) - (0.9477 \times 13.08) = 0.7594\ (a_{yx})$					

Note: n = 1 to N; Σx_n = sum of all x-values from 1 to N; Σy_n = sum of all y-values from 1 to N

The slope that is developed when NIRS-predicted data are compared with reference data can be used to correct deviations between the NIRS and reference data in the prediction set (and subsequent samples). This slope correction is carried out by multiplying all of the calibration constants including the intercept, by b. Following that, a is added to (or subtracted from) the original calibration intercept. This will change the size of errors at extremes of the reference data. This will not change the value of r, which means that it will not have changed the closeness of the relationship between the x_n and y_n-data (n = 1 to N). Most NIRS software packages in use today include options to change slopes or biases at the touch of a key, so the operator does not have to worry about application of the actual formula. Figure 4.3 illustrates the slope and intercept. Sometimes users confuse the intercept with the bias. The intercept is a separate statistic. It can be substantially different from the bias, and should not be confused with it. An intercept can occur even if the bias is ±0.00.

Slope changes to NIR calibrations are generally not recommended. If the slope is within 0.05 of 1.00, a slope adjustment will not improve the data very much. On the other hand, in practical terms if the slope differs from 1.00, e.g. by ±0.85 to ±1.15 or by an even greater deviation, this means that the calibration will be affected. The data will apparently be improved by slope adjustment, but the model may not be reliable for prediction of future sample sets. What can happen when NIRS is applied to a problematical assignment, such as prediction of, e.g. mycotoxin in a grain or prediction of soil mineraliseable nitrogen content, is that an acceptable r and SEP may be achieved for the calibration/test set combination – the slope will however be much higher or lower than 1.00. If this happens, when the calibration model is used to predict the same substance in a new set of samples, the results will be untrustworthy, particularly at extremes of concentration.

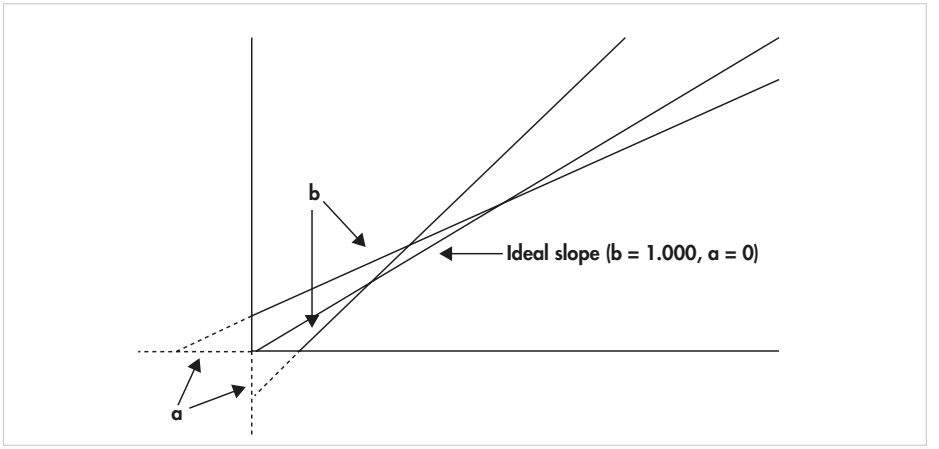


Figure 4.3 Illustrating the slope (b) and intercept (a).

Addition of predicted samples of the same type to the calibration may appear to improve the statistics, but application of the updated model to further samples may again show little improvement over the original statistical data. The main reason why this happens is that the samples that have been predicted and added to the original calibration set in an attempt to improve the calibration, are also subject to the same factors that caused the original poor slope. Because of this they are unlikely to change the slope, so that it is logical that they will not bring much improvement to the application. The reasons for the poor slope have to be researched, and perhaps NIRS is not suitable for that application.

An example of the use of b to correct results is given in Table 4.13. The values are typical of those obtained in the NIRS analysis of wheat for protein content, but the principle of the table is applicable to any application. The predicted values

of y_n (the reference data and n = 1 to N) are commonly referred to as y-hat ($\bar{y}$). Y-hat is the symbol that represents the predicted equation for a line of best fit in linear regression. The equation is $\bar{y} = a + bx$, where b is the slope and a is the intercept. It is used to differentiate between the predicted data and the observed data (y). The y-hat-value is also used in calculating the residuals of ($y_n - \bar{y}$), where n = 1 to N. The residuals are the differences between the observed and predicted values. The example illustrates the application of the regression equation to predict the values of y_n (n = 1 to N). In practice, validation of the calibration model, and determination of the slope and bias should be based on far more than 10 samples. The RPD statistic (see section 4.1.13) in Table 4.13 relates the SEP to the SD of the reference data in the sample set used for validation. Notice that the pattern of differences between reference and NIRS protein content differed after application of the slope/bias correction. The root mean square error of prediction (RMSEP) was higher before application of the correction, but the same as the SEP after correction. The SEP was slightly improved by the correction. The correction did not affect the correlation coefficient, but b became 1.0000 and a 0.0000.

Table 4.13 Application of regression coefficient (slope) (b) and bias correction

Observation / Statistic	Protein (%) Reference (1)	Protein (%) NIRS (2)	Difference* (1 – 2)	Corrected protein NIRS (3)	Difference** (1 – 3)
	12.3	12.0	0.3	12.18	0.12
	10.9	10.7	0.2	10.82	0.08
	13.6	13.2	0.4	13.43	0.17
	15.2	15.1	0.1	15.41	-0.21
	10.3	10.1	0.2	10.20	0.10
	13.1	13.2	-0.1	13.43	-0.33
	17.4	16.9	0.5	17.29	0.11
	9.8	9.7	0.1	9.78	0.02
	14.7	14.1	0.6	14.37	0.33
	11.9	12.1	-0.2	12.28	-0.38
Mean (%)	12.92	12.71	0.21	12.92	0.00
SD	2.381	2.272	0.251	2.370	0.231
CV (%)	18.43	17.88		18.34	
R	0.9953			0.9953	
r^2	0.9906			0.9906	

Table 4.13 Application of regression coefficient (slope) (b) and bias correction (continued)

Observation / Statistic	Protein (%) Reference (1)	Protein (%) NIRS (2)	Difference* (1 – 2)	Corrected protein NIRS (3)	Difference** (1 – 3)
b_{yx}	1.04302			1.0000	
a_{yx}	-0.3368			0.0000	
SEP (%)* **			0.251		0.231
RMSEP			0.334		0.231
RPD			9.49		10.26

Corrected protein = (NIRS Protein × b_{yx}) + a_{yx}; *Differences between reference and NIRS protein; ** Differences between reference and NIRS corrected protein; ***SD of differences between reference and NIRS protein content results

4.1.8 Distribution of differences between NIRS and reference results

The pattern, or the distribution of differences between NIRS and reference results, should be studied for any calibration involving materials that have not been tested before since the relationship between reference and NIRS results may change at higher or lower concentration. Nine different patterns may characterise the slope of the relationship between NIRS predicted and reference data. These are listed in Table 4.14. A slope change may be accompanied by a positive or negative overall bias in which a slope/bias correction may be applicable. Biases (types 8 and 9) can also occur, but with no slope difference. Figure 4.4 shows examples of slope types 1, 4, and 7. Types 8 and 9 can be corrected with a bias adjustment. All of the other types of bias can be corrected with a slope/bias adjustment.

Table 4.14 Patterns of slopes indicating relationship between reference and NIRS predicted results at low and high reference values

Observation	Low reference value	High reference value
1	Accurate	Accurate
2	Accurate	High
3	Accurate	Low
4	High	Low
5	Low	High
6	High	Accurate
7	Low	Accurate
8	High	High
9	Low	Low

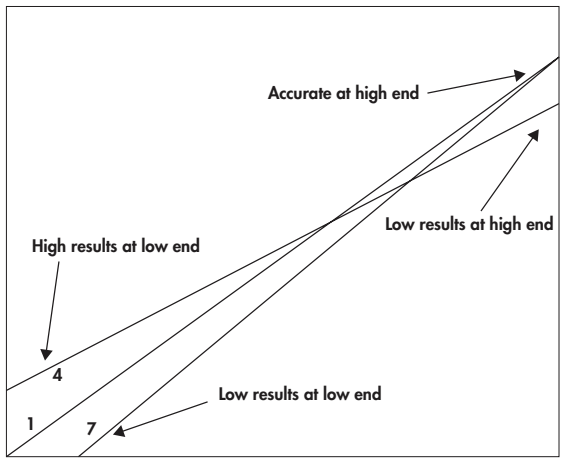


Figure 4.4 Examples of slopes with slope patterns of numbers 1, 4 and 7 from Table 4.14.

4.1.9 Standard error of a single test (SET)

The SET (sometimes referred to as the standard error of the laboratory, SEL) is the SD of the results of analysis of a check sample, whether by reference or NIRS methods, using all steps of the test, including sub-sampling and sample preparation. The sample should be tested at least 8–10 times and the SD calculated. This is discussed in more detail in Chapters 3 and 9. The SET is an intrinsic part of the SEP (see section 4.1.10). Before attempting the development of a calibration, it is essential to determine the precision (standard error) of the reference test that will be used to develop the calibration. This can be done by means of 'blind' duplicates, or by using a 'secret' check sample that is introduced under a coded number at intervals in the daily work-load (see Chapter 3, section 3.2).

To protect against hidden biases, all results for agricultural commodities should be reported on a constant moisture basis. This involves testing the sample for moisture content before and after sample preparation. For example, an error of 0.5% moisture, if unnoticed, results in an automatic bias of about 0.1% in protein-testing of cereal grains, and greater errors in the case of testing high protein material, such as soybean meal. The error in protein-testing of fresh forages or any fresh material can be even greater because of the higher moisture content. The SET should also be determined for every NIR instrument in a laboratory or any operation. It is particularly important to determine the SET for methods for the prediction of functionality, as well as other methods where some subjectivity could occur.

If the SET is determined for samples that range widely in concentration, it is useful to correlate the SET with concentration by using samples of high, medium and low levels of concentration. If the SET changes as the concentration goes up or down, this may mean that the reference method is not reliable at all levels. This is important in calibration development because an increase in the error of the reference test at different composition levels will affect the development of the calibration model. Because the principle of the NIR technique derives solely from the spectra, the NIR instrument may be more reliable than the reference method at different levels of composition.

4.1.10 Standard error of prediction (SEP)

The SEP is the SD of differences between NIRS predicted and reference values after correction for bias.

The formula for its calculation is as follows:

$$SEP = \sqrt{\langle \{\Sigma(x_n - y_n)^2 - [\Sigma(x_n - y_n)]^2 / N\} / (N - 1)\rangle}$$

Where, n = 1 to N

x_n = reference data from 1 to N

y_n = NIRS predicted data from 1 to N

Unlike the RMSEP (see section 4.1.12), the SEP is independent of bias. The SEP should be computed from the results of prediction of a set of samples that have not been used in development of the calibration. This sample set is usually called the 'prediction', or more often, the 'validation' sample set. Ideally, the sample set used in the independent validation of a calibration should consist of samples of the same type that are completely unrelated to the calibration sample set. Nevertheless, often the validation samples are part of a single population from which both calibration and validation samples sets are compiled. Ideally, a test set should be used for validation of a model using totally new samples of the same commodity. If the population of samples is large (500 or more), the requirement for the validation sample set to be made up of totally different samples is less important because sufficient variance from all sources will have been incorporated into the population, and the addition of more samples of the same type will add little to the integrity of the calibration.

The ultimate test of a calibration is the application of the calibration to fresh 'real-world' samples that have entered the testing area for the first time and represent the material for which the application of NIRS is intended. As an example, when the CGC had completed the research on development of the calibrations that

they would be using in their wheat industry for the prediction of protein content, the first 'real world' sample was a 35,000-tonne cargo of grade 1 CWRS wheat at a guaranteed protein content of 13.5%, shipped to a flour mill in England.

Computing the results of prediction of the samples used in the development of the calibration gives the SEC. The SEC is the SD of differences between NIRS predicted and reference samples in the calibration sample set. The calculations are based on a bias of 0.0 and a slope of 1.0. The spectral signals at adjacent wavelengths are highly correlated with each other, and if multiple linear regression (MLR) is used, the r and SEC statistics will progressively improve as more terms are added. This is called over-fitting of data (multicollinearity) and can be misleading. If the validation exercise indicates that r is unacceptably low and the SEP is unacceptably high, the calibration set should be predicted and the individual data viewed. The SEC may indicate the presence of one or more gross outliers, the removal of which may bring about a significant improvement in the actual r and SEP values when the validation set is predicted. Validation using a separate set of samples enables the operator to optimise the number of constants to use in either MLR equations or PLS regressions.

In practice, the SEP may not always be higher than the SEC, particularly if calibration and validation sample sets are compiled from small populations (100 samples or less), or if a 'repeatability file' of samples is included in calibration development (WinISI software) with differences in temperature or moisture contents for the purpose of adding variance to improve the stability of the calibration. The precision of the NIR instrument is often superior to that of the reference method. Sample selection for calibration and validation sets may result in one or more of the calibration samples having a higher reference test error than any of those in the validation set, which will also cause the SEP to be slightly lower than the SEC.

A note about outliers: A spectral 'outlier' is defined as a sample that does not conform to the bulk of the population in terms of the spectral data (see Chapter 5, section 5.2.11). In a validation exercise, an outlier is a sample, of which the predicted result differs from the reference result by three times the SEP or more. In the industrial world there are no outliers, everything has to be tested. However, during calibration development, samples/spectra that appear as outliers can occur. It is important to determine whether these are really outliers. NIR instruments are very precise, so that spectral outliers can be confirmed as outliers by re-scanning the sample. If the spectral data of the second scan is clearly different than that of the first scan, the sample should be re-scanned in order to determine which of the two scans were correct. If the spectral data of the second scan remain essentially

the same as for the first scan, the sample is spectrally different from the rest of the population. The source should then be traced to find what has caused it to be different, e.g. growing location or a change in ingredients or processing conditions. Additional samples of the same type should be sought and added to the calibration/validation sample sets. If a spectral outlier can be traced to a sample that is not likely to occur again (e.g. a single sample of soft wheat that has accidentally become mixed with the hard wheat), it can be removed.

The spectral characteristics of every individual sample exerts an influence on the calibration model to a certain extent. This is called 'leverage'. The degree to which most samples exert leverage is related to their chemical, physical and physicochemical make-up which determines their positions within the PCs. Most of the samples exert approximately the same amount of leverage. Their position along the PCs of the variance governs the sign and degree of their influence.

Outliers interact with the incident energy in a way that does not conform to the rest of the population. As a result, they exert a much higher degree of leverage than they should and tend to bias the calibration model toward themselves to an extent that the model may not predict future samples that do conform to the population, accurately. Figure 4.5 shows the influence of an extreme outlier on a calibration.

Outliers can occur where the predicted results differ considerably from the reference results, but the spectral data of the outliers look normal. These may be 'reference' or 'chemical' outliers. Sometimes these are due to entry errors, e.g. entry of a reference value of 1.43% protein instead of 14.3%. The authenticity of suspected reference outliers can be verified by repeating the reference test. Again, if the new result differs from the original result significantly (e.g. by 2–3 times the SEL), the reference test must be repeated in order to determine which test was correct. If the new result brings the sample in line with the rest of the population, the sample is no longer an outlier. If the new result agrees to the previous result to within the SET, the sample will persist as an outlier. If this is the case, a decision has to be made.

For calibration purposes, up to 2% outliers can be retained in a large population without distracting from the integrity of the calibration. They represent samples of which the spectral data differ from the mean of the population, and will add stability to the calibration because such samples will happen in the day-to-day industrial operation. If there is a much higher proportion of apparent outliers, they are probably not outliers. It may mean that the reference data are unreliable,

the calibration requires some fine-tuning, e.g. by optimising mathematical pretreatment or wavelength range, or that NIRS is not applicable to the prediction of the parameter sought. One anomaly of NIRS that has been with us ever since the introduction of the computerised spectrophotometer, concerns spectra that 'look' normal yet predict as outliers. One explanation of this is the influence of external factors such as growing location on the spectra of samples of the same chemical composition (see Chapter 9, section 9.1.4).

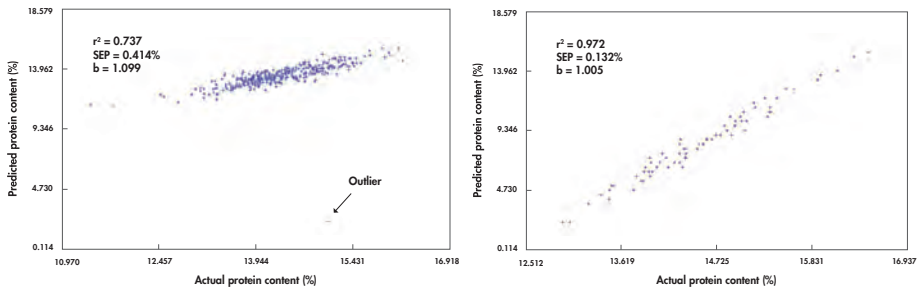


Figure 4.5 Model evaluation before and after removal of a gross outlier.

A word of caution when using PLS regression in developing calibrations is that, although the use of all available wavelengths may lead the operator to believe that over-fitting is impossible, very attractive values for r and SEC for up to 15 PLS factors do not necessarily mean that those values will be achieved for the validation set. A type of over-fitting can occur if too many factors are used. The r, r^2 and SEP values when a validation set is predicted can indicate that a different optimum number of PLS factors should be used from that determined and suggested by the software. The number of PLS factors should be optimised by prediction in the same way as the number of wavelength points in MLR. This can be done by means of cross-validation (see section 5.2.9). Both over-fitting and under-fitting can occur with PLS regression. In applications where only a single PLS factor appears to have been used, the calibration is usually unreliable. Because of all of the variance introduced by the spectra and reference data it is essentially impossible for a single factor to account for all of the variance in a data set.

Theoretically most of the variance in an NIRS PLS calibration system should be accounted for by the first 3 or 4 factors so that the PLS calibration model will contain at the most 5 factors. In practice, unless the relationship between spectra and reference data is very strong, successful PLS models rarely use less than 6–8 factors. On the other hand, a useful guideline is that the lower the number of PLS factors, the better, and a PLS calibration that indicates the use of 15 factors should be regarded with caution.

The SEP should not be regarded as an indication of accuracy by itself. It is the SD of differences between reference and NIRS data, and reporting the effectiveness of an NIRS calibration model must include any bias as well as the SEP. Freedom from bias is the true indication of accuracy. Table 4.15 gives reference data with two sets of NIRS data, showing that excellent statistics for r, r^2 and SEP can be obtained with or without bias.

Table 4.15 NIRS results with and without bias

Statistic	Reference	NIRS I	Difference I	NIRS II	Difference II
	11.7	12.0	0.3	12.0	0.3
	13.9	13.8	-0.1	13.8	-0.1
	15.3	15.6	0.3	15.0	-0.3
	14.6	14.8	0.2	14.4	-0.2
	10.4	10.7	0.3	10.4	0.0
	12.3	12.6	0.3	12.5	0.2
	13.2	13.3	0.1	13.3	0.1
	16.1	16.3	0.2	15.5	-0.6
	14.7	14.8	0.1	14.5	-0.2
	11.2	11.4	0.2	11.4	0.2
	13.4	13.5	0.1	13.5	0.1
	12.8	13.0	0.2	13.0	0.2
	14.0	14.4	0.4	13.9	-0.1
	12.5	12.8	0.3	12.7	0.2
	15.7	16.0	0.3	15.3	-0.4
Mean	13.45	13.67	0.213	13.41	-0.04
SD	1.67	1.66		1.45	
SEP*			0.125*		0.261*
Bias		0.22		-0.04	
R		0.997		0.995	
r^2		0.994		0.991	
b_{yx}		0.9897		0.8674	
a_{yx}		0.35		1.74	
RPD			13.4		6.4

*SEP = standard deviation of differences

Note that, while the values of r^2 are excellent for both sets of NIRS data the SEP, b- and a-values for the NIRS II data are quite different from that of NIRS I data. Table 4.16 shows the effects of predicting the results using the b- and a-values from the NIRS I and II data. The predicted results using the NIRS I data show a consistent bias of about 0.2% which can be easily adjusted. Prediction of the results using the regression formula from the NIRS II data give results typical of a type 4 slope (see section 4.1.8 and Figure 4.4).

Table 4.16 Prediction results using regression equations NIRS I and NIRS II from Table 4.15

Actual results	Prediction I*	Prediction II*
10	10.25	10.42
11	11.24	11.29
12	12.23	12.15
13	13.22	13.02
14	14.21	13.89
15	15.20	14.75
16	16.19	15.62

4.1.11 Standard error of cross-validation (SECV)

Cross-validation legitimately uses the same samples for validation as it uses for calibration development. It does this by leaving out one sample, or groups of a few samples, and developing the calibration model with the remaining samples. The sample, or group of samples, that has been eliminated is predicted and the residual(s) recorded. This sample(s) is then replaced, another sample or group of samples is eliminated and the process repeated until all samples have been used in model development and predicted with the residuals recorded. None of the samples predicted have actually been used in development of the model with which they have been predicted. If only one sample has been left out at a time the method is called 'full' or 'leave-one-out' cross-validation. If batches of samples have been left out it is called 'segmented' or 'group' cross-validation. The standard deviation of differences of the residuals between NIRS predicted and reference data is the SECV.

Cross-validation has become widely-accepted and is used in many publications. It is useful to determine the optimum number of PLS factors for a calibration. It has the drawback that full information is not generated for bias or slope/intercept

because all of the samples have been drawn from the same population. For small sample sets (up to N = 100), cross-validation is preferred over the test-set method for evaluation of a calibration model. The reason is that with small sample sets there is a predisposition to select samples that represent as full a range as possible in the test set. Prediction of these tends to 'flatter' the calibration model. If there are only 100 or so samples sorted from high to low in terms of reference data, and the 80:20 system is used for setting up calibration and validation sample sets, there will be only about 20 samples in the test (validation) set. With cross-validation, particularly leave-one-out cross-validation, the validation will be based on the whole population of 100 or so samples, which is better. Leave-one-out cross-validation only takes a few seconds with modern computers and gives a more realistic picture of the efficiency of the calibration model for these small sample sets.

If cross-validation is to be used for evaluation of calibrations for large sample sets, e.g. 300 or more it is best to use segmented (group) cross-validation, leaving out groups of 5 or more samples. With the random system for selection of the groups, this adds the variance within each group to the overall cross-validation. Note that when random selection of the 'left-out' groups is used, the r^2 and SECV statistics can be expected to differ slightly if the exercise is repeated, because the composition of the groups will have differed due to the random selection.

With large sample sets, e.g. between 300 and 400 or more, a good system is to sort the samples by the values of the constituent, then use about 20% of the samples as a first validation test set, another 20% for a second validation test set and to use the remaining 60% to develop the calibration model. The first validation test set (the 'optimisation' set) is used to optimise wavelength range and spectral data pretreatment. This can be done using group cross-validation to save time. The first test-set is then added into the calibration model and the updated model re-evaluated using the second test-set. The second validation set (the 'validation' set) will then have been developed from 80% of the population and is used to evaluate the resulting most updated model.

4.1.12 Root mean square error of prediction (RMSEP)

The RMSEP, also termed root mean square of the differences (RMSD), or simply the root mean square error (RMSE) gives another measure of the efficiency of a calibration equation. It is the root mean square of differences between NIRS predicted and reference results. The RMSEP includes bias error. If no bias exists the SEP and RMSEP or RMSD are identical.

The formula for computing the RMSEP is:

$$RMSEP = \sqrt{\langle [\Sigma(x_n - y_n)^2] / N \rangle}$$

Where, n = 1 to N

The RMSEP (or RMSE) does not give information on the magnitude or sign of the bias and in reporting the statistics of an NIRS calibration the use of the SEP together with the bias is preferred. The use of the RMSEP statistic, rather than the SEP is preferred by some workers. The data in Table 4.17 show how the RMSEP is related to the SEP. In the case of the original NIRS data, where there was a positive bias of 0.21, the RMSEP was substantially higher than the SEP. After correction using the regression equation where there was no bias, the RMSEP and SEP values were equal. The separate SEP and bias figures are preferred for reporting accuracy since the RMSEP by itself incorporates the bias but does not indicate its size or sign. This is immaterial if the bias, or both $\bar{x}$ and $\bar{y}$, are reported separately. Table 4.17 gives details of calculation of the SEP and the RMSEP.

Table 4.17 Details for calculating the SEP and RMSEP

Observation	Reference protein (%) (x)	NIRS predicted protein (%) (y)	Difference (y − x)
1	13.28	13.38	0.10
2	13.15	13.60	0.45
3	13.64	14.04	0.40
4	13.58	13.56	-0.02
5	13.42	13.64	0.22
6	13.61	13.81	0.20
7	13.47	13.59	0.12
8	13.33	13.43	0.10
9	13.67	13.66	-0.01
10	13.29	13.60	0.31
11	13.59	13.84	0.25
12	3.38	13.53	0.15
Sum (Σx_n)	161.41	163.64	2.27
Mean ($\bar{x}$)	13.45	13.64	0.189
SD	0.168	0.182	0.147
Bias	0.14		

Table 4.17 Details for calculating the SEP and RMSEP (continued)

Observation	Reference protein (%) (x)	NIRS predicted protein (%) (y)	Difference (y − x)
$\sqrt{\langle\{\Sigma(x_n - y_n)^2 - [(\Sigma(x_n - y_n))^2/N]\}/(N-1)\rangle}$ $\sqrt{\langle\{0.6669 - [2.27/12]\}/11\rangle}$ $\sqrt{\langle\{0.6669 - 0.4294\}/11\rangle}$		= SEP (SD of differences = SEP $= \sqrt{\langle 0.2375/11\rangle}$ = 0.147	
$\sqrt{\langle[\Sigma(x_n - y_n)^2]/N\rangle}$ $\sqrt{\langle 0.6669/12\rangle}$		= RMSEP (root mean square of differences) = 0.236%	

Note: n = 1 to N; Σx_n = sum of all x-values from 1 to N; Σy_n = sum of all y-values from 1 to N

4.1.13 Ratio of SEP to the SD (RPD)

The RPD is a non-dimensional statistic for the evaluation of a NIRS calibration model. The Ratio of (standard error of) Prediction to (standard) Deviation, or RPD, is a simple statistic that enables the relative evaluation of a SEP in terms of the SD of the reference data of the validation sample set. It is calculated by dividing the SD of the reference values used in the validation (SD_y), by the SEP:

$$RPD = SD_y / SEP$$

An alternative formula for computing the RPD is:

$$RPD = 1/\sqrt{(1 - r^2)}$$

Most people find SD_y / SEP easier to remember, but the $1/\sqrt{(1 - r^2)}$ formula gives a useful estimate of the RPD when information on the SD of the reference data is not available. The RPD can be computed for evaluations for cross-validation as well as for test-set evaluation. This is useful when only cross-validation is used in calibration development.

One of the most frequently-used statistics for evaluating the efficiency of NIRS analysis (as determined by the size and consistency of deviations from reference analyses) is the SEP. If the SEP value is equal to the SD_y (it may be even higher), it means that the instrument is not predicting the reference values at all because the differences between each predicted value and the mean are the same as the differences between the reference values and the mean. If this happens the RPD will be 1.0. The SEP should be considerably lower than the SD_y, and ideally the ratio of the SD_y to SEP should be 3 or higher.

Using the data in Table 4.13, the original RPD was calculated as 2.38 / 0.251 = 9.5. Because the SEP was reduced after slope/bias correction, the RPD changed to

10.3. For a series of fairly uniform samples where the SD is not high, the RPD may not be as high as this, but should indicate that the SEP is appreciably lower than the SD. For example, RPD-values of even 2.5–3.0 will verify accurate analysis if the SD is only 0.4–0.5. From a practical point of view, if the SD is very low for a population of reasonable size (60 or more), this may indicate that the variance is so low that frequent analysis is not necessary. This type of situation may appear, e.g. in a flour mill, and in many industrial applications where the objective is to maintain uniformity in the product and minimise variance. Despite the low range in reference values, and consequent low SD, there is still a need for regular analysis and quality control in order to ensure that specifications are being met. Table 4.18 gives the interpretation of some values of the RPD for grains, flours and meals, for composition.

Table 4.18 The RPD statistic – grains, flours, constituents

RPD-value	Classification	Application
0.0–2.3	Very poor	Not recommended
2.4–3.0	Poor	Very rough screening
3.1–4.9	Fair	Screening
5.0–6.4	Good	Quality control
6.5–8.0	Very good	Process control
8.1+	Excellent	Any application

NIRS has been successfully applied to a very variable range of commodities, constituents and functionality factors. Because of complications with the inherent heterogeneity of the samples, sample preparation, sample presentation to the instrument and difficulties with reference testing, RPD-values of as high as 3.0 may be difficult to achieve with some applications. Even with lower values for the RPD, this statistic still gives a useful indication of the evaluation of applications of NIRS for these types of industrial and scientific analytical work. Table 4.19 gives the interpretation of RPD-values for some different and more 'difficult' applications, such as analysis of forages, high moisture materials, such as silages vegetables, fruits, meat and fish, soils and manures, where the application of NIRS is affected by the more complex nature of the materials.

Table 4.19 is also applicable to the prediction of functionality factors such as grain texture, digestibility and energy or other parameters that are not associated with any 'classical' absorbers. The different interpretation does not change the statistical consequences of the lower values, but it places them in a category of

explanation that is more realistic in terms of the results likely to be achieved with these applications.

Table 4.19 The RPD statistic – forages, feeds, functionality parameters

RPD-value	Classification	Application
0.0–1.9	Very poor	Not recommended
2.0–2.4	Poor	Rough screening
2.5–2.9	Fair	Screening
3.0–3.4	Good	Quality control
3.5–4.0	Very good	Process control
4.1+	Excellent	Any application

4.2 Summary of statistical terms

To summarise the interpretation of the statistical terms, the SEP indicates the variability in deviations of X from Y, and the bias shows the average amount by which the results differ. The SEP, together with the bias, describe the overall accuracy of the test procedure. The RPD relates the SEP to the SD of the reference data used in validation data, and standardises the interpretation of the SEP. High values for the RPD (ideally 5 or more, but at least 3) indicate efficient NIRS predictions. Remember that the SEP is a standard deviation, and as such is subject to the same limitations as the SD (see Table 4.4).

The r or r^2 indicates the closeness of fit between the NIRS and reference data over the range of composition. A high r with a low SEP and bias, together with a slope close to 1.0, means that the NIRS results are accurate over the anticipated range, and likely to remain so provided that these statistics were based on a sufficient number of observations. Useful guidelines are to use at least 25 samples for every wavelength, using MLR, and at least 15 samples per factor using PLS regression. A high r (>0.95), with a bias throughout, or at one or both extremes of composition, means that these discrepancies will almost certainly persist.

Qualitative and quantitative analysis by NIRS can be compared in this way – with high values of r^2 and RPD (0.95+ and 8+) quantitative analytical results can be expected that would be good enough for quality control in a factory, and qualitative results in terms of classification that would identify classes correctly more than 90% of the time. With values in the 'screening' area (e.g. an RPD of 2.5–2.9, Table 4.19) quantitative analysis results could be expected that would estimate composition or functionality well enough to identify high, medium and

low levels, and qualitative results that would identify classes correctly about 66–70% of the time, again probably still better than methods used before NIRS, and better than doing no testing at all.

The slope shows the degree to which NIRS predicted values change, relative to reference values. A slope of near 1.0 is excellent because it shows that the rate of change in both sets of data is identical. Deviations from 1.0 of greater than 0.05 may appear to require slope, and possibly bias, correction. Deviations of greater than 0.1 are more significant and require investigation as to the cause. The regression statistics 'b' and 'a' can be used in the equation y = a + bx to correct the slope and bias. In practice, slope changes are not recommended and are rarely used. Based on test set evaluation, if the slope is different from 1.0 by more than ±0.10 the calibration model will probably not work well for subsequent samples.

Slope/bias adjustment will improve the accuracy of prediction at extremes of reference data. The correlation coefficient will not change. If the correlation coefficient is low (<0.8), it is usually not possible to obtain consistently high accuracy by NIRS analysis even after slope/bias adjustment because the calibration model will not be reliable. A low correlation coefficient between NIRS and reference data means that the NIRS analysis has not been successful. If all sources of error for poor NIRS results have been carefully studied, including reference analysis, and no improvement can be achieved, the sad truth may be that NIRS is not applicable to the analysis.

The coefficient of determination, r^2, the bias and the RPD are the most meaningful statistics for 'instant' appraisal of analytical efficiency by NIRS and comparison of calibrations. The slope is an indication of the potential effectiveness of the calibration.

4.3 Reporting of NIRS results

To present the results of an application of NIRS effectively, up to 18 pieces of information should be included. These are summarised in Table 4.20. In many cases the r^2 and SEP are reported without all of the information that had to be involved in the development of the calibration and its application. The use of cross-validation provides no information on the slope or bias. The slope is most important when NIRS is applied to complex materials, such as fresh or mature forages, soils, manures, composts, meat and fish, fresh fruit and vegetables and others. Reporting the intercept is optional, and its inclusion illustrates differences between the intercept and the bias. More items are discussed in a Journal of Near infrared Spectroscopy tutorial (Williams, Dardenne & Flinn, 2017).

Table 4.20 Statistical and other information to be included when reporting the results of an NIRS application

Number	Statistical term/other information
1	Source of samples
2	Number of samples in calibration and test sets
3	Sample preparation and storage method
4	Reference method
5	Standard error of the reference method on the material to be tested (SET or SEL)
6	Mean of reference data
7	SD of reference data
8	Method of developing calibration model, e.g. MLR or PLS
9	Number of wavelengths (MLR) or number factors (PLS) used in final model
10	Cross-validation system (e.g. leave-one-out or size of blocks left out)
11	SECV or RMSECV
12	SEP or RMSEP
13	r and/or r^2
14	Regression coefficient or slope
15	Intercept (optional)
16	Bias
17	RPD
18	Standard error of the NIRS method

Provided that the calibration has been developed using samples that have represented all of the variance of the population, the standard error of the NIRS method may be superior to that of the reference method. This is because there are usually more sources of error in the reference method.

Reference

Williams, P, Dardenne, P & Flinn, P. 2017. Tutorial: Items to be included in a report on a near infrared spectroscopy project. *Journal of Near Infrared Spectroscopy*, 25(2), 85–90. https://doi.org/10.1177/0967033517702395

05

Introduction to NIRS software

This chapter describes the purpose of NIRS software which is to translate spectral data into the information required on items such as composition, functionality and other properties of biological samples. This is usually achieved by stepwise MLR, principal component regression (PCR) and PLS regression, and sometimes by ANN (artificial neural networks) regression.

An ANN calibration is based on thousands of spectra, all of which are accompanied by reference data. These have been assembled by instrument companies (originally by Tecator) from their clients using spectra recorded on a number of instruments with reference data from an equal number of laboratories which incorporate a wide range of operating conditions with respect to temperature, relative humidity and dust conditions. Provided that the results of the reference testing are unbiased, the combination of spectral and reference data from a large number of instruments and sources will add valuable variance in spectral data with the same range of reference data to the calibration model. This has the overall effect of stabilising the model. ANN calibration models demand very large numbers of samples, and it is difficult to obtain sufficient samples with reliable reference data for constituents such as, e.g. NDF. It is even more difficult to acquire sufficient

samples for development of an ANN calibration for things like metabolisable energy, where each test costs C$800.00 or more per sample.

Modern computers and software can process data very quickly. In the early days of computerised spectrophotometers, development of a four-wavelength point calibration model took several hours using MLR. This limited research on optimisation of wavelength ranges and mathematical pretreatment of the spectral data. Modern computers can process the same calibration model in a few seconds. This allows the operator to optimise the wavelength range and mathematical pretreatments (see section 5.2.5) in a very short time. A detailed study of these features that would have taken several weeks 45 years ago can now be completed in a couple of hours. Operators are advised to learn the software thoroughly, and then to test variations in wavelength ranges, mathematical pretreatment of spectral data and scatter correction in detail before deciding which is the most reliable calibration model. This is a very good investment in time and allows operators to become skilled in the use of the software.

5.1 Types of NIRS software

Modern NIRS literature and conference papers contain many references to new forms of chemometrics that are aimed at resolving all the problems of calibration. The talented people that design and programme software, such as WinISI, Pirouette®, OPUS, NIRCal and The Unscrambler® read these papers and attend the conferences. If they note a new feature that may be effective they test it themselves in their own laboratories, and if it does indeed prove to be useful they will incorporate it into the software.

Software can be grouped into two main classes, dedicated and generic software. The first type is dedicated to specific NIR instruments. It enables recording of spectra, operation of the instrument in regular analytical mode and instrument diagnostics. Calibration models can be developed directly using the dedicated software, and the models are automatically set into the instrument for use in future analysis. The FOSS WinISI, Metrohm Vision, BÜCHI NIRCal and the Bruker OPUS are examples of dedicated NIRS software.

Generic software, such as MATLAB®, The Unscrambler®, Grams®, PLS_Toolbox® and Solo® is usually not limited to use in the NIR wavelength range and may be used to process other than spectral data. Instruments that require generic software for calibration development and related functions are provided with dedicated software that allows recording of spectral data, sample identification, instrument diagnostics, acceptance of the calibration models and routine analysis. For

development of calibration models, the spectral data must be transferred to the generic software to be used in calibration development and evaluation, usually on another computer, and the final model transposed back into the instrument for use in future analysis.

The terms 'gap' and 'segment' were originally introduced by Neotec/Pacific Scientific to explain the dimensions of the distance between two wavelength points when a derivative was created. The gap provides the dimensions of the derivative, while the segment describes the degree of smoothing (see section 5.2.5).

5.2 Purposes of NIRS software

The purposes of NIRS software include:

1. instrument set-up and standardisation;
2. recording of spectra;
3. regular ('day-to-day') analysis; and
4. data handling:
   a. viewing (plotting) of spectra, singly or grouped;
   b. entering and editing sample ID;
   c. entering and editing reference data;
   d. capability of selecting spectra from large sets of data files, such as the individual files recorded by, e.g. OPUS, for use in setting up calibration and validation files;
   e. capability to import data and spectra from clipboard;
   f. averaging of replicate spectra;
   g. file set-up and organisation, e.g. for calibration and validation sample sets, and development of repeatibility files;
   h. capability to add or remove samples to existing files individually or in blocks;
   i. capability to delete and restore samples;
   j. capability to create new files by removing spectra from other files and combining them;
   k. compatibility with other software, such as Excel for importing reference data;
   l. capability to 'zoom' on areas of spectra to magnify them;
   m. capability to subtract spectra from one another; and
   n. movement of the cursor, to any point of the spectrum to identify wavelength and absorbance.
5. mathematical pre-treatment of optical data (smoothing and derivative development). The software should allow mathematical pretreatment and display of selected spectra from spectra files of any size;

6. application of scatter correction;
7. optimisation of wavelength range;
8. development of calibration models;
9. evaluation of calibration models (statistical analysis)
10. slope/bias correction;
11. outlier detection;
12. discriminant analysis (classification);
13. networking of instruments;
14. instrument diagnostics;
15. special graphics in order to display statistical parameters and to illustrate publications and reports;
16. capability to change colours of backgrounds, graphs and spectra;
17. report generation;
18. client service; and
19. capability of clients to extend factory calibrations with their own samples.

Comprehensive software for use in application of NIR technology using computerised spectrophotometers should possess all these options. To describe each of the software packages in detail is beyond the scope of this book. The manuals that accompany the instruments describe how the software works and its capabilities. Software packages present features, such as file manipulation, data-processing, scatter correction, instrument diagnostics and graphics in their own unique manner. The algorithms (methods) that they use to develop derivatives, principal components and PLS factors may differ among software companies, but the end results are quite similar. The manuals that accompany the software usually do not include all of the steps entailed in using all of the options. Users are advised to take the time to explore all of the options, and to develop their own stepwise operating manuals for training purposes.

5.2.1 Instrument set-up and standardisation

Some instrument companies have developed software to set up and standardise instruments to a master instrument at company headquarters so that their clients can use the same calibrations for all of their instruments for the same applications. The clients send in a sample of the commodity with which they are going to work. The company makes a sealed cell, scans it on their master instrument and sends the cell back to the client, together with a diskette that contains an equation. The client scans the cell and applies the equation. This standardises the instrument to the instrument company's master instrument. Calibrations can then be transferred among all instruments that have been standardised. On

receipt of the calibrations by the client they should be checked to verify that no biases have been introduced.

5.2.2 Recording of spectra

Recording of spectra involves creation of spectral data files. The process includes setting up projects under which files are to be stored, naming files in which spectra are to be stored, number of scans per sample, specification of whether the data are to be recorded in analytical or calibration development modes, identifying method of sample naming and numbering, identifying method of adding reference data, specifying the resolution (of interferometers) and other features specific to the individual instrument or software. Furthermore, the method of transferring the spectral data to generic software may be specified. Some instruments do not allow the operator to add reference data at the time of recording spectra. These data are added after scanning using dedicated or generic software.

Some instrument/software combinations create a separate file for each sample and feature a system for identifying spectra/samples for data-processing, as well as calibration model development. Some software programmes allow reference data to be imported directly from Excel files. It is imperative that when this very useful and time-saving feature is used, that the spectra and Excel reference data are sorted to correspond with each other, and check samples, duplicates are removed before importing the reference data.

Most NIRS reflectance instruments record spectral data as the log 1/R where 'R' is apparent reflectance. NIT instruments record data as the log 1/T where 'T' is apparent transmittance. The terms absorbance or optical density (OD) are often used instead of log 1/R. A higher log 1/R value means that more of the incident radiation has been absorbed by the sample. Because more of the radiation is absorbed, less is reflected, hence 1/R gives a higher value. This affects the vertical position of the spectra, the more the degree of absorbance, the higher will the spectra appear on a display (see Chapter 1, Figure 1.3).

5.2.3 Routine analysis

Routine analysis, including on-line analysis, means 'real world' analysis at commodity receiving and shipping points, strategic locations in processing plants and so on. Software can enable the instrument to be set up to scan the samples, report the data and store the spectra as required. In processing operations, the data can be recorded automatically and displayed so that quality control can be monitored and maintained, and adjustments made if necessary. Selected spectra

can subsequently be used to update and extend calibrations. The spectra can be stored in separate files if the operator selects samples for this purpose during daily analysis.

5.2.4 Data handling

Data handling includes viewing the spectra, adding and editing sample identification (ID) and reference data, as well as setting up files for calibration development and evaluation. Most stand-alone reflectance instruments allow the operator to view the numerical spectral data, usually as log 1/R or absorbance values. An experienced operator can detect obvious abnormalities in spectral data. In these cases, the numerical data can only be displayed graphically as spectra in the software used for calibration development. Operators of scanning spectrometers sometimes need to view spectra singly and the associated software should include that option. Some software allows the spectral data to be exported to Excel as vertical columns. This can be useful, e.g. if special graphics are to be used in preparing reports or scientific papers.

Based on scans on a scanning spectrophotometer, the log 1/R values in the 1100–2500 nm range for commodities such as wheat flour vary from around 0.01 to 0.70. For ground wheat-meal, such as produced by the Cyclone grinder with a 1.00 mm screen, log 1/R values range from about 0.01 to 0.75. There is not much difference in the spectral data between flour and ground wheat. The significant differences in mean particle size between flour and wheat meal are balanced by the fact that the wheat meal is much darker than the flour, which affects reflectance from the surface. Log 1/R values for whole wheat kernels range from about 0.11 to 1.15 for hard wheats and from 0.20 to 1.25 for soft wheats. Values for durum wheats vary from 0.29 to 1.3. Slightly different values are likely to be obtained with different NIRS reflectance instruments. Values of spectral data recorded in reflectance mode on whole grains by interferometers may be higher, and optical densities range up to 1.4. Absorbance of other materials, such as high moisture samples can reach over 2.0 units. These spectra should be treated with caution since some non-linearity could occur that will interfere with calibration development. Interferometers record spectral data in the form of cm^{-1} (reciprocal entimetres).

The absorbance values for NIT instruments usually vary from about 1.5 to 4.0 for grains. A value of 4.5 is very high, and adjustment should be made to reduce these values to around 3.0 to 3.5. This can be achieved by changing the path length (actually sample thickness; see Chapter 6, sections 6.2.1 & 6.2.2).

Sample ID of individual samples is usually added before or during scanning. Some instruments limit ID to 8 characters. It is a good idea to include some numerals in the sample ID, so that the spectra can be sorted by sample number for subsequent importing of reference data, e.g. from Excel. Reference data can be imported to several systems directly from Excel. This is particularly useful when large databases are available for calibration development, and several constituents/parameters have to be added. Data have to be entered into Excel manually, usually in the analytical chemistry laboratory at the time of reference testing. Importing the reference data to the NIRS software saves a lot of time and avoids further typographical errors. It is essential that the reference data are imported in the same order that the spectra are recorded. This can be done by sorting both the spectra and reference data by sample ID. Software systems may differ in the way that samples are sorted numerically. The ones may all be sorted in sequence, e.g. 1–19, then all of the 100s before passing to the twos (2, 20s, 200s). Operators should check that the instrument software sorts samples in the same way as Excel. As mentioned above, check samples and blind duplicates must be removed from the files of spectra before importing from Excel, because they change the order of samples in the spectra file relative to the reference data in the Excel file. It is important to verify the order of both spectra and reference data before calibration development. Inadvertent inclusion of a duplicate spectrum will displace all of the reference data which will result in spurious calibration statistics.

With a population of samples much larger than 100, the test-set system is often preferred to cross-validation for evaluation of the calibration model. Samples can be sorted by the constituent for which the calibration is being developed. The recommended system for large populations of up to 500 samples is to set up two files, the first contains 20% of the samples. The remaining 80% forms the calibration set. Sorting on the basis of the constituent, and taking every fifth sample of the sorted samples, ensures that each file has similar distributions of samples on the basis of the constituent. The file with 80% of the samples is the calibration file. It is used to develop numerous trial calibration models representing various spectral data pretreatments and wavelength ranges using cross-validation. Selected models are then used to predict the validation sample set in order to determine the best of the trial calibration models. This is termed optimisation of wavelength range and mathematical pretreatment. This best model is the one with the highest r^2, lowest SEP (standard error of prediction), lowest SET (standard error of a single test) and highest RPD (ratio of the SEP to the SD_y).

For bigger populations the system suggested in Chapter 4 (see section 4.1.11) is preferred where the population is sorted then sub-divided into two validation

sets, each 20% with 60% of the samples being used for the development of the calibration model. The model is optimised using the first validation set, extended by adding this first set of 20% of the samples into the calibration set and updating the new calibration set which now contains most of the samples. The calibration is finally evaluated using the second set of 20% of the original samples, which have not been used previously. To optimise mathematical pretreatment, gaps and segments can be set at, e.g. 2, 4, 8, 10 and 20 wavelength points, and cross-validation is recommended for this. This means a total of 25 possible combinations of gap and segment (smoothing). These are all tested by prediction of the first validation sample set. With a given commodity, certain combinations of gap and segment will emerge as optimum, as wavelength ranges are tested. This means that not all gap and segment combinations will need to be tested. This apparently formidable project can be completed in about an hour.

Cross-validation is recommended for evaluation of any calibration models based on small sample sets, e.g. up to 100–150 samples. The best model from cross-validation again has the highest r^2, lowest SECV (standard error of cross-validation) and highest RPD. If cross-validation is to be used, there is no need to sort or subdivide the samples, because all of the samples will be used.

All of the options listed in this section are useful and should be available in the software. Features, such as moving the cursor, 'zooming' in on areas of the spectra to examine them in more detail and subtraction of spectra to identify areas of maximum difference are useful in finding out more about the spectra. Averaging of spectra with subsequent display is useful to see at a glance whether two types or classes of material differ from one another spectrally.

5.2.5 Mathematical pretreatment of optical data

Calibration of the original discrete 6-filter instruments involved MLR of the 'raw' log 1/R data against the reference data to obtain b and a. Modern scanning spectrometers and interferometers use a variety of pretreatments of the spectral data, the most common of which is smoothing, multiplicative scatter correction and derivatisation. Although the concepts of smoothing and derivative development had been discussed by earlier workers, the initiative of using a 'derivative' of the log 1/R data was introduced by Karl Norris (Williams & Norris, 1984), for use with computerised spectrometers.

Many monochromators record log 1/R at intervals of 2 nm called wavelength points. Calibrations can often be improved by 'smoothing' the spectral data. Derivatives are usually developed on smoothed data. This can be done in more than one

way, but the most popular way is called 'moving point' or 'boxcar' smoothing. This is done by averaging the spectral data over a number of wavelength points, specified by the operator, and replacing each log 1/R value with the average. For example, if the spectral data were to be smoothed by 4 wavelength points, the spectral data would be handled in sets of five points. The mean would be calculated by summing the two points on either side of the central point and dividing by 4. The mean would replace the log 1/R value of the central wavelength point. Each wavelength point would then represent the mean of four data points. This would be repeated throughout the spectrum. The few wavelength points at each end would then not be available.

The number of 2 nm wavelength points used for averaging has been referred to as the 'segment'. The bigger the segment, the more smoothing has been done. Too much smoothing can cause loss of some resolution, and potentially useful data may be lost. The effect of smoothing is most obvious in areas of the spectrum where the system noise is highest. This is illustrated in Table 5.1, which also shows the effect of wavelength range on the smoothing. The effect of smoothing is determined here by calculating the SD of the spectral data from 10 spectra before and after smoothing. Smoothing over 10 wavelength points had a greater effect than smoothing over 5 points in reducing the SD of the spectral (log 1/R) data. The effect of smoothing is clearly indicated in Figure 5.1, where the smoothed spectrum in the area around 2300 nm shows how the changes in the spectra are evened out by the smoothing action.

Table 5.1 Influence of smoothing on spectral data (log 1/R) of wheat flour at three wavelength ranges

λ	416–450 nm			868–910 nm			2316–2346 nm		
	Raw	**S5***	**S10***	**Raw**	**S5**	**S10**	**Raw**	**S5**	**S10**
Mean	0.1816	0.1814	0.1811	0.11352	0.11353	0.11356	0.6150	0.6151	0.6156
SD	0.0128	0.0126	0.0116	0.00292	0.00290	0.00280	0.00572	0.00563	0.00535
CV	7.071	6.929	6.424	2.570	2.551	2.469	0.9302	0.9154	0.8694
Diff.**	-	-2.01%	-9.15%	-	-0.74%	-3.93%	-	-1.59%	-6.54%

* S5 = raw log 1/R data smoothed by 5 points; S10 = raw log 1/R data smoothed by 10 points;
** Diff = difference in CV (%) from raw data induced by smoothing; CV = coefficient of variation

The wavelength region over which the derivative is developed is the 'gap'. For example, the gap could be set at 10 wavelength points and the segment (smoothing) at 5 points. The derivative could then be defined as having a gap of 10 and a segment of 5 (i.e. 2 10 5 1). Such a derivative would extend over

10 wavelength points in the spectrum regardless of the degree of smoothing. Third and fourth derivatives are sometimes mentioned in scientific papers, but only first and second derivatives are routinely calculated and used. The use of a derivative of the log 1/R is intended to correct for baseline drift and for some of the variance in diffuse reflectance caused by particle characteristics.

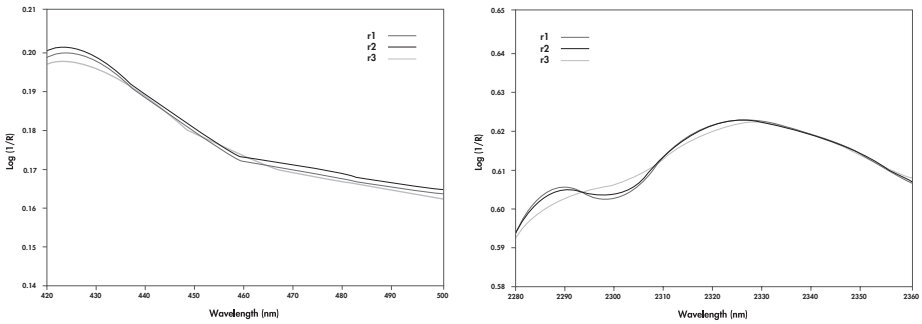


Figure 5.1 Effect of smoothing between 420–500 nm (left) and 2280–2360 nm on flour spectra. The straightest spectrum in each plot has been smoothed over 10 wavelength points.

Developing a derivative of the log 1/R brings all of the spectra to a common baseline. The second derivative assists in resolution of overlapping absorption bands for viewing purposes. Figure 5.2 illustrates the principle of derivatives. The wavelength point A is the log 1/R value. Points B and C are the wavelength points used in computing the derivatives. The first derivative of OD or absorbance, uses two wavelength points, A and B. These may have already been smoothed. The log 1/R values for wavelength point B are subtracted from the data of point A, and replace the smoothed log 1/R data for the first wavelength point. This is continued throughout the spectral range. The distance in wavelength points between points A and B is the 'gap'.

Using WinISI software, a first derivative with gap and segment each of 4 points can be written as 1 4 4 1 where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of smoothing data points and the fourth is the second smoothing (a value of 1 indicates no smoothing thus no effect). The formula for calculating the first derivative is:

$$1^{st} \text{ derivative} = \log 1/R_A - \log 1/R_B$$

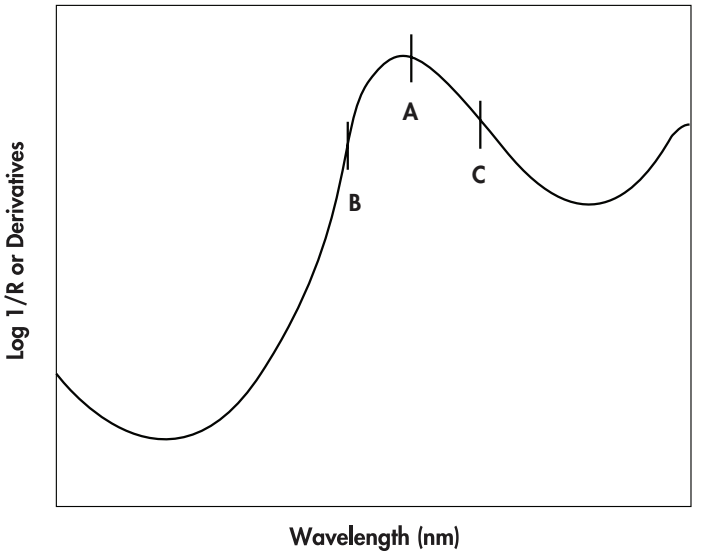


Figure 5.2 Illustration of the principle of derivatives. Point A is the log 1/R. Points B and C are the wavelength points used in computing the derivatives.

Development of a derivative involves subtraction of values of spectral data so that the individual data for derivatised spectra will be considerably smaller than the original spectral data. For example, the actual data from whole grain at 1410 nm will change from about 0.65 for raw data to about 0.02 for first derivative data. The same applies to second derivative data (Figure 5.6).

The 2nd derivative uses three wavelength points. Twice the values of the data at point B are subtracted from the data at point A, and the data from the third point C are added to the result. The formula is:

$$2^{nd} \text{ derivative} = \log 1/R_A - (2 \times \log 1/R_B) + \log 1/R_C$$

The 2nd derivative has the effect of sharpening the absorption bands, which appear as sharp 'valleys'. This makes it easier to identify the bands where absorptions overlap. Figure 5.6 illustrates the 2nd derivative of whole wheat in the NIR region. The combination water band at 1908 nm and its overtones at 1424, 1202 and 968 nm are very clear. Remember that all fundamental absorptions occur in the mid-IR region, and the bands that appear in the NIR region are all overtones or combinations of these. The combination bands in the NIR region (above 1850 nm) are first overtones of the combinations that arise in the mid-IR region.

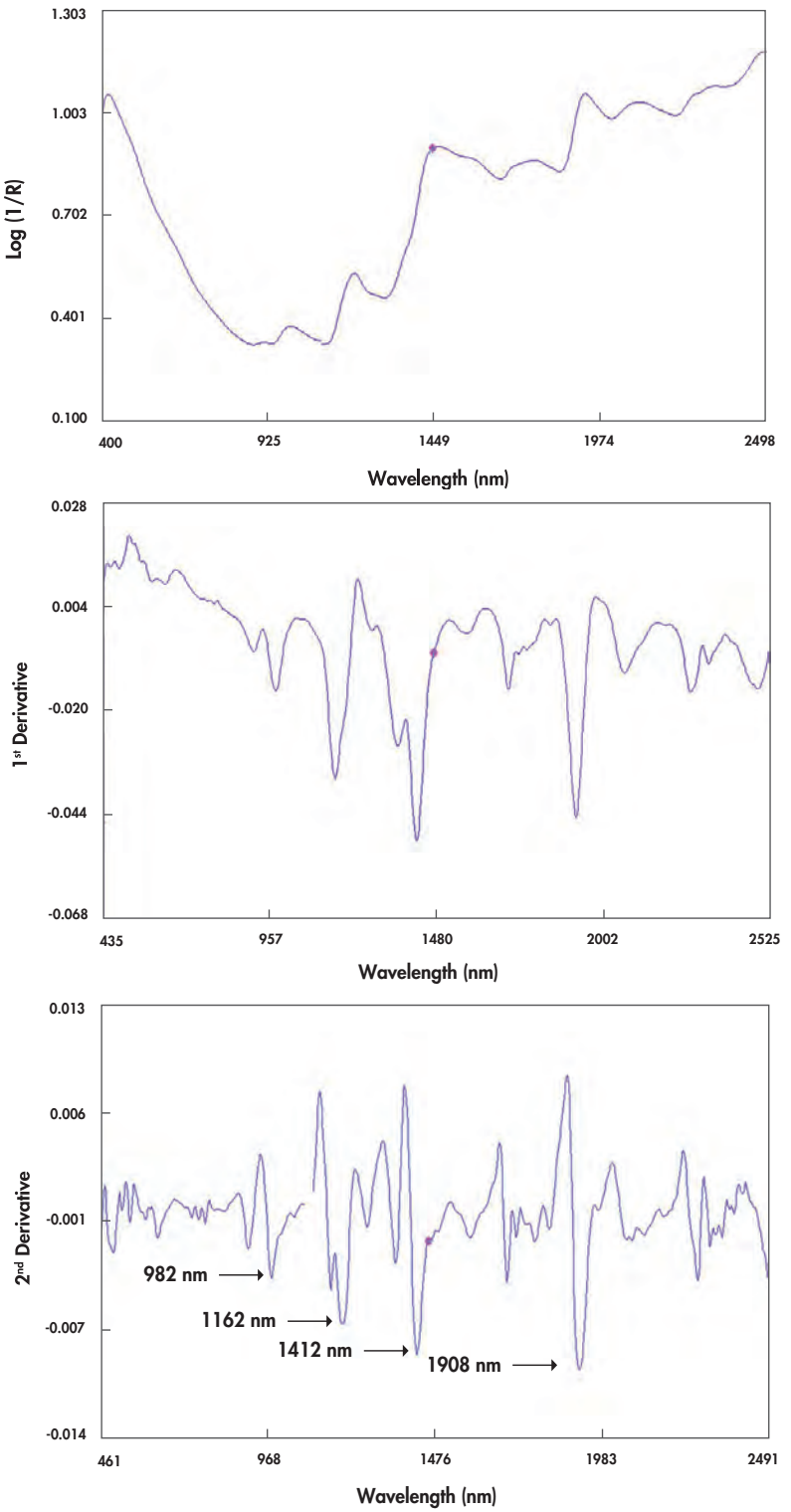


Figure 5.3 Log 1/R spectrum of whole wheat and the influence of the 1st derivative (1 4 4 1) and 2nd derivative (2 4 4 1) on the NIR spectrum, with moisture bands indicated.

Models based on 1st derivatives often give better results than 2nd derivatives, but the spectra are not so easy to interpret. Second derivatives orientate bands downwards, and the minima ('valleys') are useful in identifying what is likely to be causing the main absorption band. Fourth derivatives provide an even sharper definition of absorption bands, and the bands are orientated upwards. Higher derivatives above the second derivative are not favoured for calibration model development, mainly because the signal-to-noise (S/N) ratio is reduced at higher derivatives. The S/N ration is defined as the 'ratio of the useable NIR spectral data received by the detector to the amount of signal contributed by extraneous factors (the noise) to the total signal (amount of information) received by the detector', as reflected from or transmitted through the sample. The signal itself carries the relevant NIR information. 'Noise' is defined as disturbance in the electronic signal that interferes with the required utilisation of the signal. It can be caused by several factors, some of which may not be related to the log 1/R or derivatised log 1/R information. The higher the S/N ratio, the better.

Many absorption bands in the NIR region overlap each other and cannot be clearly seen. One effect of the 2nd derivative is to make absorption bands appear sharper, because it removes most of this overlapping. For calibration development the smoothing, if log 1/R is to be used, and the sizes of the gap and segment (smoothing) if a derivative is to be used, should be optimised. An excellent explanation of derivatives is given by Hruschka (2001).

Instruments such as the Perten Model 7200 and Zeiss Corona, both diode array instruments, record spectral data at intervals of between 5 and 6 nm. The spectra appear to be much smoother than monochromator spectra, but the larger wavelength interval has an effect similar to smoothing 2 nm interval spectra by up to 3 wavelength points. These instruments tend to work quite well using their raw log 1/R spectral data, and derivatives may not be as helpful as they are with grating monochromator instruments.

5.2.6 Application of scatter correction

In NIR technology light scattering means that the pathway of light through a sample, or reflected from the sample is deflected from a straight path to the detector by features in the material being scanned. Scatter is always a factor in NIRS applications because of the physical nature of the materials to which it is being applied. The concept of multiplicative scatter correction (MSC) was introduced by Harald Martens and co-workers in 1983 (Geladi, MacDougal & Martens, 1985; Martens & Næs, 2001). The option is available in most software

packages used for NIRS calibration. There is more than one version of scatter correction available, including standard normal variate (SNV) with or without de-trend (Barnes, Dhanoa & Lister, 1989), but standard MSC appears to be the most suitable.

The process involves correction for differences between the individual spectral data for all samples at a wavelength point, and the average spectral data for all of the samples at that wavelength. Figure 5.4 illustrates the effect of application of MSC to a set of wheat spectra selected from a particle size index (kernel texture) sample set.

The variance in spectral data is significantly reduced, but this does not necessarily mean that the calibration will be improved. The normal scattering of the light energy signal by the sample is part of the variance needed for computing and stabilising, the equation. Application of scatter correction only improves the calibration if the multiplicative and/or additive effects of the scattering interfere with the natural scattering of the energy caused by the particle and composition characteristics of the sample set. The basis of successful NIRS calibration is the translation of inherent variance in the spectral data into reliable information of composition and functionality, by regression of this spectral variance against the reference data. Removal of some of this variance may inhibit, rather than enhance the integrity of the calibration.

Figure 5.4 shows the different effect of applying the SNV and de-trend version of scatter correction to the same sample set. Note the differences in the shape of the spectra, especially in the low wavelength range. The data in Table 5.2 show the effects of MSC, SNV and de-trend and only SNV on the prediction of NDF by the crucible method in barley and kernel texture in wheat.

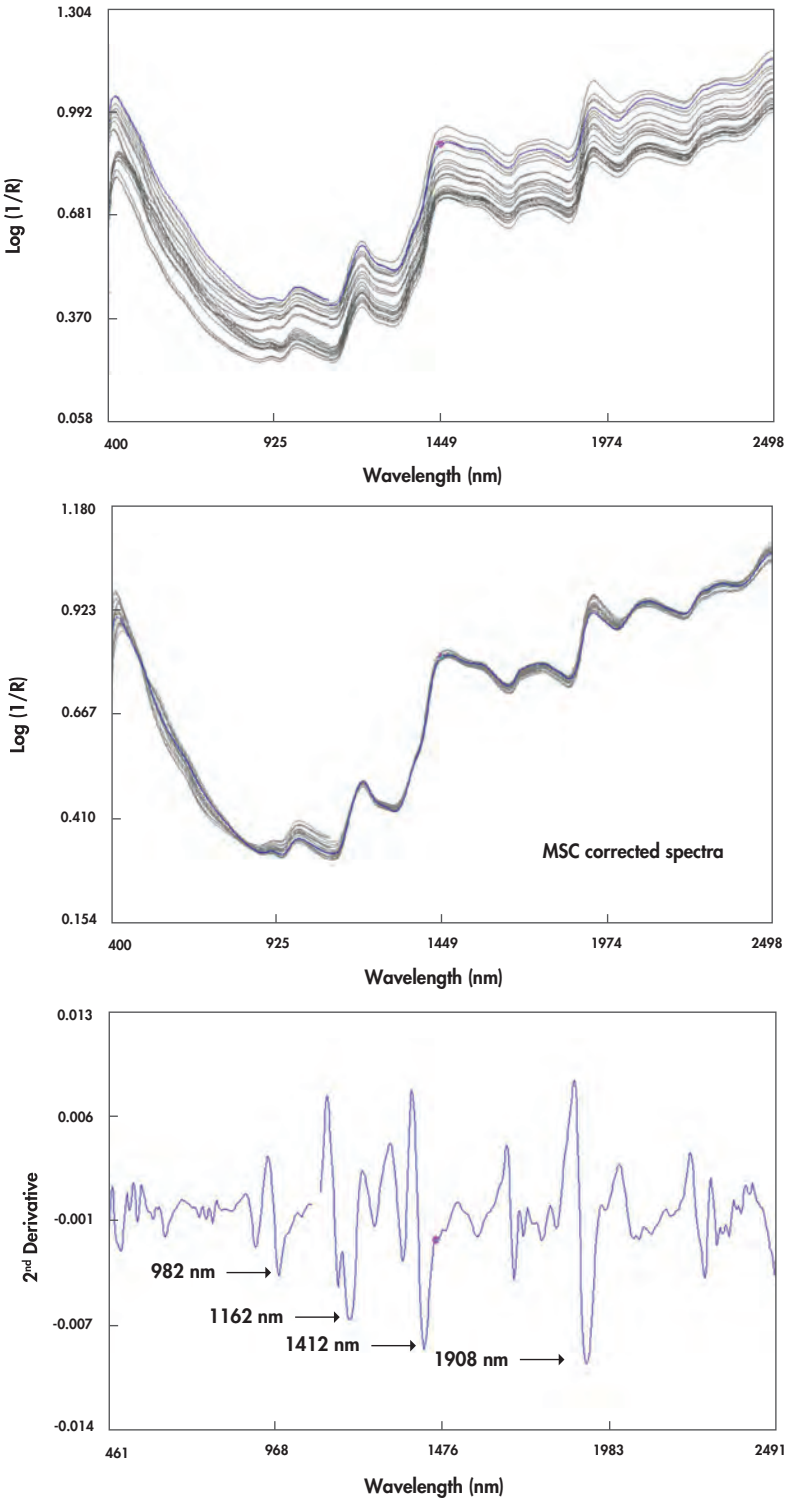


Figure 5.4 Raw log 1/R spectral of whole wheat of different kernel texture and the effect of applying MSC, SNV and de-trend to the raw spectral data.

Table 5.2 Effect of applying three versions of multiplicative scatter correction to log 1/R spectra to the prediction of two parameters

Mathematical pretreatment	r^2	SEP	Bias	Slope	RPD
Neutral detergent fibre – barley					
No scatter correction	0.964	0.646	0.136	1.019	5.25
Multiplicative scatter correction (MSC)	0.967	0.615	0.132	1.003	5.51
Standard normal variate (SNV)	0.960	0.685	0.200	1.027	4.95
SNV and de-trending	0.965	0.648	0.187	1.035	5.23
Kernel texture – wheat					
No scatter correction	0.892	1.938	0.070	1.027	3.03
MSC	0.887	1.975	0.057	1.012	2.98
SNV	0.881	2.029	0.101	1.021	2.90
SNV and de-trending	0.871	2.118	0.154	1.107	2.78

In only one case did the application of MSC or SNV with or without de-trending lead to an improvement in the results, and the effects of scatter correction were not of great consequence in either of these applications. In using NIRS for prediction of whole wheat kernel texture, the scatter induced by differences in particle characteristics is essential to development of the calibration model, and application of scatter correction may inhibit model development. In applications to grains and derived products MSC is usually more successful than SNV and de-trending. SNV and de-trending was originally developed for use with applications of NIRS to forages, thus in materials such as ground roughages, straws, soil and other more complex materials, this method of scatter correction has had a more significant influence. Modern computers are very fast and the usefulness of testing the influence of scatter correction is recommended for any new application.

5.2.7 Optimisation of wavelength range

Modern computerised spectrophotometers scan the wavelength range from 400–2500 nm, while Fourier-transform interferometers (FT-NIRS instruments) scan from 12,500–4000 cm^{-1}. Calibrations or many constituent/material combinations do not necessarily require the full wavelength range, and when using dedicated or generic software to develop equations for application in analysis, the wavelength range and mathematical pretreatment of raw spectral data should be optimised. Some dedicated software, such as OPUS and NIRCal offer an optimisation option in their calibration procedure. The software scans the calibration set and gives recommendations for the best combination of wavelength and mathematical

pre-treatment. This operation is discussed in more detail in Chapter 9 (see section 9.1.12). It can be carried out using either cross-validation or test-set validation. The most robust calibrations are a combination of optimum wavelength selection and mathematical pretreatment that provide the best SEP and precision (SET). If efficiency in NIRS predictions is to be defined by the SEP then consistency in arriving at the same result repeatedly (precision) has to be an inherent part of it.

5.2.8 Development of calibration models

Stand-alone whole-grain instruments can be delivered pre-calibrated for the most commonly-tested constituents. The factory calibrations use ANN or PLS regression. These instruments have to be calibrated for other constituents and parameters. The early generation filter instruments originally used any software that could supply MLR. Some of these instruments could also be calibrated using PLS regression even though they had as few as 20 wavelength points (discrete filters).

Calibration model development can be carried out by several methods. The two methods most frequently used in calibration development for computerised spectrophotometers are MLR and PLS regression, which is the method most frequently used. In NIRS calibration model development MLR regresses all of the X-values (spectral data at up to 1050 wavelength points) against each Y-value for the population of spectra (samples). There are four forms of MLR, i.e. forward step-up, forward step-wise, backward stepwise and quotient mathematics. Forward step-up selects a wavelength based on the best correlation between spectral and reference data. This selection accounts for a certain amount of the total variance in combined spectral and reference data. The second wavelength is selected on the basis of the best correlation between the remaining spectral and optical data. The first wavelength is fixed when the second wavelength is selected, and subsequent wavelength selections are retained in the order in which they are selected. This is rarely the most successful method.

Forward stepwise MLR selects the first wavelength in the same way. When the second wavelength is selected, the software searches all of the variance to determine whether the first wavelength is best fitted to the second, in view of the now-changing pattern of variance. If it is not, the first wavelength is removed and replaced by a more suitable wavelength. This process is repeated until the specified number of wavelengths have been selected. The wavelength first selected may not even be retained in the final equation.

Forward stepwise MLR has proven itself to be the most useful form of MLR. Table 5.3 illustrates the difference between forward step-up and stepwise regression on actual sample sets for prediction of protein content in whole wheat grains (N = 107). Notice that a) the results were significantly better using stepwise regression, b) the wavelengths selected first by both step-up and stepwise regression were similar, c) after the first wavelength, wavelengths did not change very much during step-up, but changed significantly during stepwise regression, and d) that the first wavelength selected by stepwise regression did not appear in the final calibration model at all.

Table 5.3 Comparison of forward step-up and stepwise multiple linear regression (MLR) when developing models for protein content (%) in whole wheat grains

Wavelengths (nm)*	r^2	SEP (%)	Bias	Slope	RPD
Step-up MLR					
1) 2204	0.386	1.27	0.151	1.023	1.00
2) 2204, 2244	0.774	0.815	0.052	1.149	1.56
3) 2204, 2244, 2198	0.796	0.775	0.063	1.125	1.64
4) 2204, 2244, 2198, 2194	0.819	0.732	0.065	1.054	1.73
Step-wise MLR					
1) 2210	0.452	1.13	0.146	1.025	1.12
2) 1760, 1580	0.804	0.757	0.134	1.014	1.68
3) 1520, 1730, 1550	0.922	0.461	0.075	1.029	2.75
4) 1700,1550, 1430, 1520	0.986	0.206	0.014	1.041	6.16

Backward stepwise regression was the method preferred for some fixed-filter instruments. By this method, the spectral data from all of the filters are regressed against the reference data and an equation developed and evaluated. Subsequent calibration models are developed by eliminating wavelengths one by one. The models are developed and evaluated at each step. The number of wavelengths (in this case, filters) that gives the lowest SEP is accepted as the selected equation.

The fourth form of MLR is the quotient mathematics version. Here, the best wavelength is selected as in forward step-up or stepwise MLR and retained. This wavelength is designated as the numerator. A second wavelength is then selected to serve as the denominator. The system is then revisited in order to determine whether the first wavelength is indeed the best numerator to go with the selected denominator. This process, developed by Karl Norris (Norris & Williams 1984) is repeated until the most suitable numerator and denominator have been selected.

The process can be repeated to select a second pair of wavelengths (or more). The most recent version of the Norris quotient software allows optimisation of the mathematical pretreatment of both the numerator and the denominator. The equations all have to be evaluated by prediction of a validation set or by cross-validation. Of the four MLR methods, forward and backward stepwise MLR are the only two that have actually been employed in commercial software and instruments.

The spectral data at all wavelength points are highly inter-correlated throughout the spectral range although the degree of inter-correlation decreases, the further wavelength points are distant from each other. This is illustrated in Table 5.4.

Table 5.4 Correlation coefficient (r) matrix for spectral data among representative wavelength points

Wavelengths (nm)						
Wavelengths (nm)	**1154**	**1426**	**1696**	**1910**	**2054**	**2310**
982	0.9975	0.9424	0.9451	0.8998	0.8961	0.8851
1154		0.9613	0.9608	0.9251	0.9184	0.9057
1426			0.9968	0.9910	0.9877	0.9786
1696				0.9820	0.9871	0.9826
1910					0.9934	0.9838
2054						0.9976
2310						-

For any sample set, inter-correlations between spectral data at immediately adjacent wavelength points are often in excess of 0.999, but begin to fall off slightly even within 2 wavelength points. Because of the high degree of inter-correlation among spectral data, wavelength points may be selected by MLR that are very close to one another. This was apparent in Table 5.3 in the step-up regression example. This is because the first wavelength selected accounts for the highest proportion of the variance in spectral and reference data. Because spectral data of adjacent wavelengths are highly correlated with one another with step-up regression the second wavelength selected is likely to be close to the first wavelength. This does not address the residual variance effectively.

Stepwise MLR is comparable with the PLS model in that wavelengths subsequent to the first wavelength take in hand residual variance in both the spectral and reference data.

MLR suffers from the possibility of over-fitting data. This can give a misleading impression of the efficiency of an equation containing, e.g. 8 wavelength points. From a set of wavelength points selected by stepwise MLR, the optimum number of wavelengths can best be determined by validation using a separate set of samples. Computerised spectrometers access up to 1050 data points. 'Over-fitting' data, when using MLR means that, mathematically, the r^2 and SEC (standard error of calibration) values improve as more data points are added. This can be misleading unless the calibration equations are tested by validation with a test-set. Table 5.5 illustrates this for prediction of β-glucan in barley.

Table 5.5 Comparison of SEC and SEP for β-glucan in barley using increasing number of wavelength points from in MLR

Number of wavelength points	r^2	SEC	r^2	SEP
1	0.496	0.362	0.244	0.394
2	0.623	0.313	0.336	0.369
3	0.653	0.301	0.383	0.356
4	0.680	0.288	0.440	0.339
5	0.721	0.270	0.507	0.318
6	0.759	0.250	0.424	0.344
7	0.771	0.244	0.497	0.321
8	0.779	0.240	0.492	0.322

Although these results were not outstanding, the r^2 and SEC continued to improve as more wavelength points were added, whereas for prediction (validation) the r^2 and SEP did not improve after 5 wavelength points. A trap into which users may fall would be to use the SEC result without further verification, e.g. use the 8-wavelength equation directly from Table 5.5.

PLS regression was introduced to NIR technology in the early 1980s as an alternative to MLR for calibration development that would minimise the likelihood of over-fitting data (Martens & Næs, 2001). The technique of PLS regression uses all of the wavelength points in populations of up to 1050 data points. The regression equation is based on factors computed similar to PCA of the total variance in the system, however, in this case both spectral and reference data are used when calculation the factors. The PLS factors are sometimes referred to as 'latent variables'. In practice, most of the variance is accounted for by 8 factors or less. In

PLS regression, the spectral data and the constituent reference data are modeled together to obtain an optimal regression.

It is possible to over-fit PLS equations by utilising too many factors. It is also possible to under-fit. In that case, some of the useful information in the spectra is not captured. As with any other system for evaluation of calibrations, the most reliable method is to evaluate equations developed from 2 or 3 to up to 15 PLS factors. Equations are stored for the increasing number of factors, and used to predict the results of a test set. This will enable optimisation of the number of factors. Alternatively cross-validation can be used. Some software automatically stops selecting PLS factors when they reach the optimum, based on cross-validation, using the calibration sample set. When the operator is setting up these systems for calibration development the best technique is to specify a reasonably large number of PLS factors, such as 15, more than is likely to be required for the calibration. The software will stop when the optimum number has been reached.

Some software systems offer PLS1 and PLS2 options. With PLS1 the spectral data together with one dependent variable (reference data) are used to compute the PLS loadings (weights). With PLS2 the spectral data are combined with more than a single dependent variable in computing the PLS factors. When the reference data include constituents and functionality factors they are often not highly inter-correlated. For example, moisture content is usually not related to any other constituent. Unless the constituents are highly correlated with one another the addition of this extra variance in reference data rarely improves the calibration, and PLS1 is the more widely-used method.

Table 5.6 gives typical results using PLS regression on the same data set as the one used in Table 5.5, and demonstrates that the apparent improvement in r^2 and SEC as more factors were used were not matched by improvements in the r^2 and SEP achieved when a test set was used for validation. The r^2 and SEC-values gave a misleading impression of the efficiency of the calibration. This is often the case with more difficult applications. For example, for predicting protein content of wheat, the PLS SEC-values are closer to the SEP-values than they are for the prediction of β-glucan in barley. This can be verified by displaying the loadings (weights) generated during development of PLS regressions.

PLS regression does not always out-perform MLR and where possible, both methods should be tested. The wavelengths selected by PLS are often the same as, or close to those selected by MLR. These loadings can be displayed across the spectral range. What appear to be ‘peaks’ and ‘valleys’ in the loadings displays are

areas of wavelength that have been used in development of the PLS equations. Table 5.7 gives some typical results, using MLR and PLS regression for two very different examples.

Table 5.6 Comparison of SEC and SEP for β-glucan in barley using PLS regression

Number of PLS factors	Calibration		Validation	
	r^2	SEC	r^2	SEP
2	0.526	0.358	0.241	0.394
3	0.612	0.326	0.310	0.376
4	0.625	0.324	0.346	0.366
5	0.702	0.291	0.463	0.331
6	0.716	0.287	0.466	0.331
7	0.759	0.267	0.504	0.319
8	0.788	0.253	0.533	0.309
9	0.799	0.249	0.530	0.310
10	0.823	0.236	0.494	0.322
11	0.842	0.225	0.514	0.315
12	0.855	0.218	0.585	0.292
13	0.868	0.211	0.568	0.298
14	0.883	0.200	0.547	0.295
15	0.900	0.187	0.546	0.295

The underlying principle of NIR spectroscopy is that it derives exclusively from the spectra. The use of similar wavelengths by both regression techniques is very logical, because these are the wavelengths that carry the information relevant to the development of the models. One reason why PLS regression is usually more effective than MLR is that MLR uses variance at specific wavelengths, while PLS also uses variance at wavelengths both above and below, but close to the wavelengths selected by MLR. WinISI offers the alternative of MPLS (modified partial least squares) regression. PLS regression works with covariances. A covariance can be visualised as a spectrum high in an absorbance minus a spectrum low in the absorbance. MPLS regression works with correlations. Small peaks in absorbance that correlate highly to a constituent or a parameter can contribute more strongly to an MPLS than to a PLS calibration. Both MPLS and PLS regression should be tested during the development of calibration models. Table 5.8 gives several

examples of calibration models for very different applications. In all cases the statistics for the MLR and PLS predictions are similar.

Table 5.7 Comparison of wavelength areas used by MLR and PLS regression on same sample sets

True metabolisable energy – barley (units)		Flour protein (%)	
MLR SEP = 0.223	**PLS** SEP = 0.218	**MLR** SEP = 0.108	**PLS** SEP = 0.121
Wavelengths	**Wavelengths***	**Wavelengths**	**Wavelengths***
2420	2412	2020	2016
2360	2360	2420	2420
1560	1568	1980	1980
2280	2284	2080	2084
1720	1722	1460	1454
2140	2144	2100	2096
1260	1268		
2400	2394		

*'Peaks' of PLS weights

Table 5.8 Some typical comparisons of predictions using MLR and PLS regression

Constituent	Commodity	N*	λ**	r^2	SEP	RPD	Factors**	r^2	SEP	RPD
		MLR					**PLS**			
DON (ppm)	Wheat	53	8	0.73	744	1.92	10	0.73	750	1.91
FN (s)	Wheat	174	9	0.60	44.1	1.57	14	0.59	44.2	1.57
TME (units)	Barley	56	8	0.89	0.22	5.14	7	0.89	0.22	5.26
Oil (%)	Canola	52	8	0.94	0.82	4.07	8	0.92	0.95	3.49
Moisture (%)	Canola	52	2	0.98	0.46	8.76	5	0.98	0.50	8.02
Protein (%)	Wheat flour	104	6	0.99	0.108	14.3	9	0.99	0.12	12.7
Ash (%)	Wheat flour	95	8	0.82	0.024	2.34	12	0.84	0.021	2.50

*N = number of samples in prediction set; **Number of wavelengths and factors;
DON = deoxynivalenol; FN = Falling Number; TME = True metabolisable energy

5.2.9 Evaluation of calibration models

There are three main methods for evaluation of calibration models. These are cross-validation, test-set validation and prediction of completely unknown samples. Cross-validation is favoured by many NIRS users. It is particularly useful when working with small data sets (up to about 150 samples). It involves elimination of samples from the development of the equation, either in groups or individually. The sample or group of samples is then predicted using the equation that has been developed using the remaining samples, and replaced in the original sample set. A second sample or group is then removed and the exercise repeated until all samples have been used in development of the calibration and predicted, but none of them have actually been used during the computing of the equations. The software lets the operator specify the group size (number of samples per group). For the most efficient use of cross-validation in such small sample sets, single samples can be eliminated/predicted. This is called 'leave-one-out' cross-validation. With larger sample sets group, or segmented cross-validation is recommended. Samples are left out in groups of 5 or more samples. By using group cross-validation, with random selection, the variance within the group, as well as among them is used, which improves the evaluation, even groups as small as 2 or 3 samples in the group. The final result is displayed as SECV, accompanied by the r^2-value.

Cross-validation is theoretically satisfactory for the evaluation of any NIRS calibration equation. It saves the steps involved in setting up separate calibration and validation sample sets. But because cross-validation is carried out on the same overall sample set it suffers from the fact that it does not provide full information on the bias or slope. The RPD statistic can be applied as in test-set evaluation.

Cross-validation is particularly useful in optimising mathematical data pre-treatment, including scatter correction. It is also useful in detecting gross errors in calibration. The SEC is obtained by predicting the samples that have been used in the calibration. This can be useful if the SEP value is high but smaller than the SEC because, if there are really bad outliers they can be detected and removed before attempting to evaluate the calibration on 'normal' samples. Such outliers should be investigated and if the error is caused by incorrect documentation or reference data entry these can be corrected and the samples can be returned to the calibration sample set (see Chapter 4, section 4.1.10). If they are spectral outliers, the source(s) should be investigated.

With a population of samples much larger than 150, the test-set system is often preferred to cross-validation for evaluation of the calibration model. Samples

should be sorted by the constituent for which the calibration is being developed. The recommended system for large populations of up to 300 samples is to set up two files, the first contains 20% of the samples. This is the validation set. The remaining 80% forms the calibration set. Sorting on the basis of the constituent and taking every fifth sample of the sorted samples ensures that each file has similar distributions of samples on the basis of the constituent. The calibration file is used to optimise spectral data pretreatments and wavelength ranges, using cross-validation. The model with the highest r^2, lowest SEP, lowest SET (see Chapter 4, section 4.1.9), and highest RPD is used to predict the validation sample set. Scatter correction can then be applied to test whether it improves the prediction statistics.

For bigger populations the system suggested in Chapter 4 (see section 4.1.11) is preferred, where the population is sorted on the basis of reference data, then sub-divided into two validation sets, each of 20%, with 60% of the samples being used for the original development of the calibration model. The model is optimised using the first validation set using group-type cross-validation. The calibration set is then extended by adding this first set of 20% of the samples into the calibration set and up-dating the new calibration set, which now contains 80% of the samples. The calibration is finally evaluated using the second set of 20% of the original samples, which have not been used previously. To optimise mathematical pretreatment, gaps and segments can be set at, e.g. 2, 4, 8 and 10 wavelength points, and cross-validation is recommended for this. This means a total of 16 possible combinations of gap and segment (smoothing). These are all tested by prediction of the first validation sample set. With a given commodity, certain combinations of gap and segment will emerge as optimum, as wavelength ranges are tested. This means that not all gap and segment combinations will need to be tested. This apparently formidable project can be completed in about an hour.

Test-set validation involves setting up calibration (also termed 'training'), and validation sample sets. The validation set is the 'test-set', and should be composed of samples of the same general type or commodity, but assembled separately from the calibration sample set. Test-set evaluation has traditionally been achieved by using the correlation coefficient (r), the coefficient of determination (r^2), standard error of prediction SEP, the bias and the regression coefficient or slope (b) (see Chapter 4 for discussion of these statistical terms).

Test-set evaluation provides all of the statistics necessary for a true evaluation – it provides the slope (b), intercept (a) and bias data, as well as the r^2 and SEP

statistics. The RPD statistic (Williams, 2014) has been introduced as a method of relating the SEP to the SD (standard deviation) of the reference data of the samples used in validation (see section 4.1.13). It offers a simple non-dimensional method for appraisal of a calibration model, and for comparison between models. It is computed by dividing the SD of the reference data used in the validation sample set by the SEP (an operation that has to be carried out manually). It is logical to compare the variance in the differences between NIRS and reference data (SEP) to the total variance in the reference data (SD). If they are equal or nearly equal, as denoted by an RPD of 1.0, then the calibration equation is not achieving a prediction. If the test set is derived as a sub-set of the same initial sample set, cross-validation is probably as good as test set evaluation because slope or bias changes are not likely to be important. The RPD statistic can also be used to evaluate models developed using cross-validation.

Test set evaluation is best suited to large populations, and cross-validation is more suitable than test set evaluation for small populations. The reason for this is that when a test set of, e.g. 20% is removed from a population of up to 100 samples, the tendency is to select samples to represent the full range of reference data. Because the number of samples in the test set is fairly small this will tend to 'flatter' the statistics. Cross-validation uses all of the samples, and is acceptable for working with small sample sets. With very large sample sets (1000 or more), use of 20% of the population as a test set still means that the calibration will be developed from 100s of samples by either test set or cross-validation, so either method can be used.

The final and ultimate test of a calibration is to set it into the instrument and use it at-line or on-line for testing 'real world' samples. A set of these samples should be submitted for reference analysis to verify the true efficiency of the calibration as an analytical tool. Probably the first instance of this type of evaluation was the first cargo of 35,000 tonnes of wheat at guaranteed protein content of 13.5% that was shipped from Canada to a flour mill in England in 1975. Since that first shipment several 1000s of cargoes have been shipped world-wide, and no complaints were ever received by the Canadian Wheat Board, during its existence.

5.2.10 Slope and bias correction

To apply a slope and bias correction to a set of NIRS equation constants, multiply all of the constants by the slope and add the new intercept to the corrected original intercept. Most modern NIRS software provide for slope/bias correction, so the operator simply needs to apply the corrections according to the software

instructions. The slope and bias are discussed in Chapter 6. Slope correction is generally not recommended (see Chapter 6, section 6.2.20).

5.2.11 Outlier detection

Because NIRS is totally based on the spectra, anything that affects the spectra will affect the relationship between the spectral and reference data, even at the same chemical composition. An outlier is a sample that does not conform to the rest of the population. Some definitions are based on the spectral data alone, others on the results of prediction, but in the case of spectral data the consensus appears to be that an outlier is a sample/spectrum that differs from the mean of the population by 3 or more Mahalanobis distances. Oversimplified, the Mahalanobis distance is the distance of the spectral datum for an individual sample from that of the population mean at each wavelength point. In the case of prediction results, an outlier is defined as a sample that differs from the mean of the residuals by 3 or more times the SEP (see Chapter 4, section 4.1.10).

Outliers can be detected and corrective measures taken during calibration development and validation. The most important first step to take when an outlier has been identified is to verify that it is indeed an outlier. First check that all reference data has been correctly entered, and that all sample identification is correct. The next step can best be done by re-testing the sample by the appropriate reference method. Re-testing may result in a significant change in the reference data. If this happens the sample has to be re-tested again in order to determine which of the first two results is the correct one. This is where it is of great importance to know the SET (standard error of the test) for the reference method (see Chapter 4, section 4.1.9). The SET of the reference method is the basis for comparison. If the difference between the results of re-testing is within the limits of the SET, the sample is probably a spectral outlier.

A re-scan may show the new spectrum to be significantly different from the earlier scan of the same sample. Spectral data are usually very precise. However, occasionally, because of factors such as a sudden 'spike' in the electrical supply, an uncharacteristic spectrum may be recorded (the standard electrical supply to a building can fluctuate by up to ±5 %). If this is the case, the sample should be scanned again in order to verify which scan is correct. If this shows the first scan to be irregular, the sample may no longer be an outlier. If the new scan is essentially identical to the earlier scan of the same sample and the reference tests match each other, the sample is an outlier by definition. Some samples will scan as an apparently normal spectrum, but will predict with an error large enough to qualify

as an outlier. This is because the spectral characteristics of that one sample have not matched the matrix that has been established between the calibration sample set and the reference data developed during development of the calibration. How this happens may indicate that an element that affects the spectral data has not been included in the assembly of samples used in calibration development.

The outlier may be a sample from a different origin, an origin that was not considered as a source of variance during the initial assembly of samples for calibration development, and it could be the forerunner of more samples of the same type. It then behooves the operator to search for more samples from the new origin in order to improve the integrity of the calibration. In the case of grains, the influence of different growing seasons at a single location or region tends to have a more important effect on the inherent characteristics of the grain than does that of different growing locations within a season. This is why it is important for an industrial operation to monitor NIRS instrument performance in grain analysis at the beginning of each new season.

5.2.12 Discriminant analysis (classification)

Some software packages, such as The Unscrambler®, offer an NIR discriminant (classification) analysis option (see Chapter 11, section 11.12). This enables the operator to determine the degree to which samples of (known) different types differ from each other. The data can be displayed graphically and numerically (Figure 5.5, Table 5.9). Discriminant analysis is sometimes called classification analysis. Discriminant analysis is useful for identification of materials by class, and for identification of the influence of factors such as growing locations, seasons or processing conditions. It is also useful as a screening method in order to determine whether a sample belongs to the population to which it is supposed to belong. Calibration equations can be developed for NIR classification analysis (Figure 5.5; Table 5.9) and used in combination with calibrations for composition analysis, and other parameters, so that, e.g. wheat could be classified, as well as tested for constituents, such as protein and moisture contents at delivery points. Software packages such as WinISI, The Unscrambler, Grams, OPUS, and Pirouette all offer classification options.

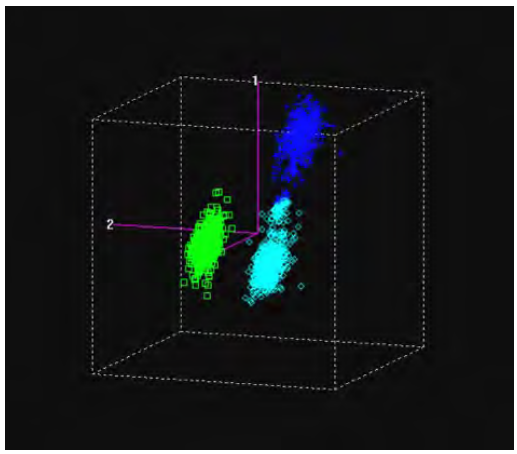


Figure 5.5 Results of NIRS classification analysis among Canadian Red Wheat classes. The cluster at the top is Canada Western Red Spring (CWRS) wheat, the group to the left CWES (Canada Western Extra Strong) and the lower right group is Canada Prairie Spring (Red) (CPSR) wheat.

Table 5.9 NIRS classification analysis among Canadian wheat classes

Wheat class*	CWRS	CWES	CPSR	% Correct classification = 93.2
Guidelines	Red	Red	Red	570–1092 nm 1 4 4 1***
Predicted (N)**	29	18	22	
Actual (N)**	28	19	22	
Wheat class	WCWW	CPSW	CSWS	% Correct classification = 99.8
Guidelines	White	White	White	570–1092 nm 0 0 4 1***
Predicted (N)	33	40	40	
Actual (N)	33	40	40	

*CWRS = Canada Western Red Spring; CWES = Canada Western Extra Strong; CPSR = Canada Prairie Spring (Red); WCWW = Western Canada White Wheat; CPSW = Canada Prairie Spring (White); CSWS = (Western) Canada Soft White Spring; **Number of samples predicted to be within the class and Number of samples actually in the class; ***Mathematical pretreatment (WinISI notation)

Qualitative analysis by NIRS is now more widely-used. In the past most applications have been quantitative, and have focused on answering the question: 'how much?' Results have been recorded in percentages, or other appropriate data. Qualitative analysis answers the questions 'what is there?' and 'will it work?' Calibrations are set up as 'teaching' sets, using samples of classes that are the most pertinent to the operation. New samples are then recognised as whether they do or do not belong, and are accordingly accepted or rejected. The teaching set has to be

very carefully prepared, and the importance of identifying and including all of the variables likely to affect the spectra of the different materials for the teaching sets are the same as for normal calibration development. Guidelines must be set to allow for a certain degree of uncertainty in use of NIRS classification.

The teaching set can be set up simply on the basis of sample type or class (e.g. different classes of wheat, as illustrated in Chapter 11 (see section 11.12), or on the basis of composition using reference data. This would approach quantitative analysis. 'Bands' of composition could be established, and new samples would be classified on the basis of to which band they belonged. For example, if an elevator intended to market wheat at two or three different protein levels, the limits of protein content would be designated for each level and the teaching set would be made up of samples with appropriate protein contents. New deliveries of wheat would then be assigned to a level. This would simplify storage. Deliveries that approached the lower limits of acceptability could be flagged by the software, and tested by normal NIRS to assign them to the correct level. The system would be applicable to any type of commodity and constituent. In plant-breeding bands could be set up for functionality factors, so that breeders could get a rough screening at early generations. There are many possible applications for qualitative NIRS. The classification model results can be expressed as classification accuracy (correct classification), but also in terms of sensitivity and specificity. Classification model sensitivity is the ability of the model to correctly identify those samples belonging to the class (true positives), whereas model specificity is the ability of the model to correctly identify those not belonging to the class (true negatives).

5.2.13 Instrument networking

Instrument networking is described in more detail in Chapter 9 (see section 9.2). The software offered by some instrument companies provides methods for identification of clients and corrections to calibration models. Some organisations develop their own systems for these procedures. The software should include a system for identification of all of the individual instruments in the network. This is usually done by means of the serial numbers of the instruments. Systems then have to be developed for monitoring of the performance of each instrument which can be done through regular supply and testing of check samples for correction of biases if necessary, and for verification that the biases have been corrected by re-scanning of the check samples.

5.2.14 Instrument diagnostics

Most modern NIRS instruments are furnished with either built-in diagnostics or a diagnostic option with the accompanying software. The instruments are fitted with an internal standard reference material, such as polystyrene, which has characteristic absorption peaks at or near 805.2, 876.5, 1143.6, 1681.3, 2166.4 and 2305.9 nm. Diagnostics monitor the performance of the instrument. System noise and wavelength accuracy (actual instrument absorbance peak as recorded by the instrument, compared with true peak) and precision (peak-to-peak noise) are the parameters usually measured by the diagnostics. The operators' manual either carries acceptable guidelines for the parameters measured, or the instrument is programmed to display error messages if specifications for the selected parameters are not achieved. Things that the operator needs to know include why the instrument fails diagnostics, e.g. what is the significance to performance if the instrument fails its diagnostics because one of the selected wavelengths has exceeded specifications by 0.2 nm. These can often be determined only by experiment.

For instruments where the diagnostics is a software option, it is advisable to run the diagnostics at least two or three times each week. The diagnostics should ideally be run at different times during the day, to determine whether performance is affected by working conditions, including changes in temperature, relative humidity, dust accumulation and so on, or by length of continuous operation. The presence of people in the room will change both the temperature and relative humidity. Instrument operating manuals usually include instructions as to what steps to take should error messages appear. Lamp failure is the most dramatic occurrence – the instrument simply stops functioning. When this happens and a new lamp is installed, it is essential that the diagnostics are run after the instrument has been turned on for an hour or two in order to check whether the performance conforms to recommended specifications. Although lamp failure is obvious, occasionally a lamp deteriorates for some time before it fails, causing anomalies in the spectra. This may be detected by the diagnostics. Failure of the cooling fan will also cause dramatic changes in performance. Accurate records of the diagnostics output are helpful to service technicians in the event of the instrument requiring repair.

Scanning a sealed check cell is also useful and can be used in conjunction with the diagnostics. The cell is prepared with the commodity or commodities that will be used most frequently and should be firmly-packed in order to prevent movement

of the material inside the cell, and sealed to prevent changes in moisture content. The sealed check cell should be scanned 10–12 times at the time of its preparation (or receipt from the instrument company), and the SD of a constituent or spectral data at selected wavelengths determined to establish the precision of scanning. It should then be scanned and the results recorded at the beginning and once or twice during each day, after the instrument has been allowed time to warm up. The daily results can be used to up-date the SD. Provided that the daily results of the check cell do not change to a degree greater than would be accommodated by the SD, there would be no need to run the diagnostics every day. This system is impossible with instruments that use hopper-type sample presentation.

5.2.15 Special graphics

This includes all aspects of displaying spectra, derivatised spectra, PLS loadings, scatter plots, discriminant analysis distributions, PCA scores plots, 3-dimensional plots and several other features. Most graphics software includes a 'zoom' option that is very useful to illustrate small differences in spectral characteristics. The software also allows for exporting or copying and pasting the figures into programmes such as Microsoft Word and PowerPoint for inclusion in reports, presentations and publications.

Graphics are usually displayed in colour and offer the options of changing the background, and ideally the changing of colour of individual spectra or data points. It is usually best to use a white background for printing. Some graphics can also be imported to other software packages that enables the change of format of objects like spectra from a coloured line to various forms of dotted lines. This is particularly useful for preparation of reports and publications. For oral presentation of data at meetings or conferences in PowerPoint – a careful consideration of colours is essential. Red and deep blue often do not project well, especially on dark backgrounds. Yellow on a white background also does not project well.

5.2.16 Report generation

Some software systems include the facility to prepare reports in a consistent format. This includes separate reports for recording the results of diagnostics, day-to-day check cell and check sample analysis. Operator records of such long-term instrument performance with regard to diagnostics and check sample analysis are useful to instrument companies when instruments require service.

5.2.17 Client service

Software used in networking includes a system for contacting participants. This is used for monitoring instrument performance, making small bias changes as required, monitoring diagnostics and detecting changes in performance among other things. Networked participants may operate many instruments all served from a central computer. Each instrument in the network is identifiable by its unique serial number.

5.2.18 Compatibility with other forms of software

This is very useful if the NIRS software that drives the instrument and is used in calibration development, is compatible with other software, such as Excel that can be used to import reference data and align it with the spectral data. Also, features such as 'autoshapes' (lines and arrows used on figures) are often useful in preparing step-wise operating manuals.

5.2.19 Capability of extending factory calibrations

Calibration transferability is becoming increasingly important as instrument companies offer factory-developed calibrations. Instrument software must include the option for clients to add samples of their own to the data-bases upon which the factory calibrations have been developed. There is no obligation for the instrument manufacturer to supply the client with the database, but the option of being able to extend the calibration will improve the usefulness of the calibration to the client. Agreement can be developed between company and client to share their spectra and reference data with the company so that both participants benefit.

References

Barnes, R.J, Dhanoa, M.S, & Lister, S.J. 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, 43, 772–777.

Geladi, P. MacDougal D & Martens, H. 1985. Linearization and scatter correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy*, 39, 491–500

Hruschka, W.R. 2001. Data analysis, wavelength selection methods. In: P. Williams & K. Norris (eds). *Near-infrared Technology in the Agriculture and Food industries*. 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. 39–58.

Martens, H. & Næs, T. 2001. Multivariate calibration by data compression. In: P. Williams & K. Norris (eds). Near-infrared Technology in the Agriculture and Food industries. 2nd Edition. St. Paul, MN: American Association of Cereal Chemists. 59–100.

Norris, K.H. & Williams, P.C. 1984. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size. *Cereal Chemistry*, 61, 158–165.

Williams, P.C. 2014. The RPD statistic: A tutorial note. *NIR News*, 25(1), 22–26.

06

NIRS Intrumentation

NIR instruments are called spectrophotometers or spectrometers, the names are interchangeable. They are driven by electricity, either from the main supply or a battery-or-generator operation. An NIR instrument is made up of a source of light energy, a system for wavelength selection, a detection system, a method for presenting the sample to the instrument, an internal standard, an internal computer with associated software to drive the instrument and record and store spectra, and a display system. At-line, benchtop instruments have all of these features. On-line instruments and computerised spectrophotometers work through an external computer, desktop or laptop. On-line, sometimes called in-line instruments are used for continuous analysis. They are also all controlled from an external desktop or laptop computer.

6.1 The basics – how do NIR instruments work?

The instruments all record information as a spectrum from the sample and from the internal standard. The internal software subtracts the spectral signal of the standard from that of the sample and uses the difference in absorbance or transmittance to compute the results. This information is translated into data

on composition and functionality of the sample by means of calibration models. Spectral data (absorbance) is recorded in the form of the log1/R for instruments that operate in reflectance mode (where R = apparent reflectance) or the log 1/T (where T = apparent transmittance) for instruments that operate in transmittance mode. The wavelengths of the absorbance data are reported in the form of nanometres (10^{-9} m). Data mathematical pretreatment, such as 1st or 2nd derivative can be applied before calibration development (see Chapter 5, section 5.2.5). Interferometers record spectral data in the form of reciprocal centimetres (cm^{-1}). A key point is that the instrument regards the calibration model as a constant factor. It will apply the model to the spectrum of any new material to which it is exposed. So, the effectiveness of any application is heavily dependent on the quality of the spectra, as well as the identification and assembly of samples that accommodate all of the variables associated with the material to be analysed.

6.2 Types of NIR Instruments

In this chapter, 'NIR' is used generically to refer to NIR reflectance instruments, operating in the wavelength area between 700–2500 nm. NIR instruments can operate in reflectance or transmittance modes. Transmittance devices are usually called NIT instruments. All NIR instruments require calibration, whether it is done by the manufacturer or the operator. Calibrations are developed by relating the spectral data of each individual sample of a calibration sample set to the reference data on composition or functionality of the sample using regression analysis (see Chapter 9). Reference data are obtained by analysis of the sample using certified or approved chemical or physiochemical (sometimes even arbitrary) methods. In reflectance instruments the internal standard is usually a plastic, such as Spectralon® (www.labsphere.com) or a ceramic tile. In transmittance instruments the reference standard is air.

The effectiveness of a calibration model is assessed by statistical methods, described in detail in Chapter 4. The accuracy is summarised by the SEP (standard error of prediction), which is the SD (standard deviation) of differences between the reference and NIRS predicted data, and the bias, while r^2 gives the overall degree to which the NIRS data agree with the reference data.

6.2.1 Reflectance vs. transmittance instruments

Reflectance instruments are those in which the light source and the detector are on the same side of the scanned sample (either above or below), and the energy is diffusely reflected back to the detector. In instruments that use a revolving

turntable where the sample is scanned from below the detector(s) are located below the sample. In some instruments the sample on the revolving turntable is irradiated from above and the detectors are located above the sample. In transmittance instruments the light from the light source is diffused through, and received by the detector on the other side of the sample. Figure 6.1 illustrates the basic differences between reflectance and transmittance instruments.

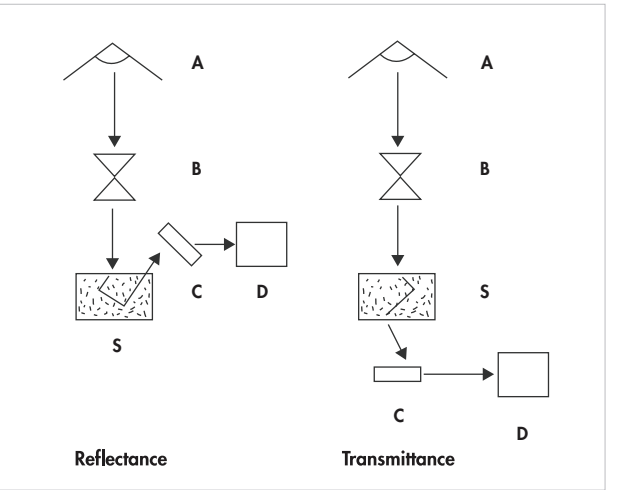


Figure 6.1 Basic instrument design. A = light source; B = wavelength selection system; C = detector; D = computer; S = sample, including sample presentation system.

Path length in terms of diffuse reflectance and transmittance spectroscopy, is the distance that radiation travels *within* the sample before it reaches the detector(s), and is important in both reflectance and transmittance spectroscopy. The path length is affected by the nature of the sample in terms of physical state, which changes the degree of scatter. This can vary from flour or ground forages with fine particle size, to coarse materials, such as chopped forages, to slurries and solid materials.

In working with transmittance instruments, the sample thickness affects the path length and has to be optimised during calibration development. This can be determined by developing and evaluating preliminary calibrations with 80–100 samples at different instrument path lengths, which can be adjusted using the instrument software. This calls for extra work, but is extremely important to obtaining the best analytical performance from a transmittance instrument with different materials. For example, using the Foss/Tecator Infratec a path length of 18 mm is optimum for analysis of whole grain wheat and barley, whereas 30 mm is optimum for maize (corn) and only 6 mm for canola seed because of its smaller seed size, and commensurate difference in sample packing density.

6.2.2 *Sample thickness and density*

Sample thickness in reflectance instruments depends partly on the thickness of the sample cell. In instruments that present the sample using a sample cell on a revolving turntable, the sample thickness has to be sufficient in order to prevent the irradiating energy from passing right through the sample, which would mean that the signals that reach the detector would not truly represent the sample. The sample thickness depends on the particle characteristics of the material. For example, a thickness of about 3 cm is sufficient for whole grain wheat, but a thickness of at least 5 cm is recommended for whole kernel maize, whereas a 3 mm thickness will be sufficient for wheat flour. Figure 6.2 shows the difference between reflectance and transmittance instruments in terms of optical geometry and sample thickness.

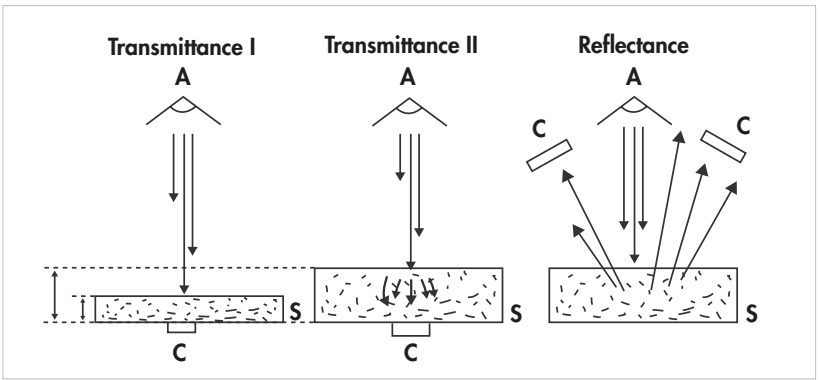


Figure 6.2 Illustrating differences in sample thickness between transmittance and reflectance instruments. A = light source; C = detector; S = sample.

Particles in suspension in liquids and slurries also affect light scattering. Imagine a clear solution of liquid. Light energy will pass straight through the liquid and will reach a detector located on the other side, and the path length, the distance that the light must travel before reaching the detector, will be the thickness of the liquid. If there are particles suspended in the liquid, e.g. from a single drop of milk, they will scatter the path of the light in random directions, so the path length will be increased, but the thickness of the liquid will not have been changed. In terms of spectroscopic analysis, the light will have collected information from the particles, but the intensity of light energy that reaches the detector from the sample will have been reduced by the amount that has been absorbed during its passage. The instrument measures the difference between the light energy applied to the sample, and the energy that reaches the detector, and translates that difference into composition or functionality data, using appropriate calibrations.

Sample density has an impact on both NIR reflectance and transmittance analysis. In transmittance the density may be as such to prevent any transmittance. In reflectance there will always be a certain amount transmitted, regardless of sample size and density (packing). In interactance (see section 6.2.3) the surface density influences the distance that the signal can penetrate the sample surface.

In diffuse reflectance from solid materials, light energy is diffusely reflected from within the sample before it reaches the detector. The particles of the sample will cause some of the light energy to be scattered, and some to be absorbed by molecular groups at the surface of and within the particles. The degree of scatter is related to the size and shape of the particles. This is greater with large, than with small particles, so that with large particles, such as whole seeds, pellets, or chopped forages, the net effect is that the energy that reaches the detector by diffuse reflectance will be reduced to a greater or lesser degree, depending on the seed, or forage fragment size. The light energy that does reach the detector will have gained information during its passage through the sample, and can be expected to contain the same type of information regardless of particle size, but less actual light energy, and therefore less actual information will reach the detector after having been more diffusely reflected from large, than from small particles. This energy is the actual information that is converted by the instrument/ calibration model into information on composition or functionality.

In the case of whole wheat and flour (Figure 6.3), the intensity of the energy that reaches the detector from the whole kernels is much less than that which reaches the detector from the flour. The degree of scatter caused by the flour particles is less than is the case with the whole kernels, but the smaller particles will have a far greater surface area. Because of the greater number of particles of the flour, the amount of useful information concerning composition will be correspondingly greater, so that in general the efficiency of diffuse reflectance analysis can be expected to be higher in flours than is the case with whole grains, and with smaller, rather than larger particles in ground grains. In the case of wheat, this is further complicated by the fact that whole grains differ in composition from the outer bran layers to the inner endosperm, whereas, in the case of wheat flour the particles consist only of endosperm, and are more uniform in composition. In ground grains the composition will be more uniformly distributed among the particles than from whole kernels. Because of this, the greater number of particles from which information can be gathered by the light energy, which represents far more kernels than are scanned in whole-grain analysis. The efficiency of NIRS analysis of ground grains can also be expected to be better than that of whole grains.

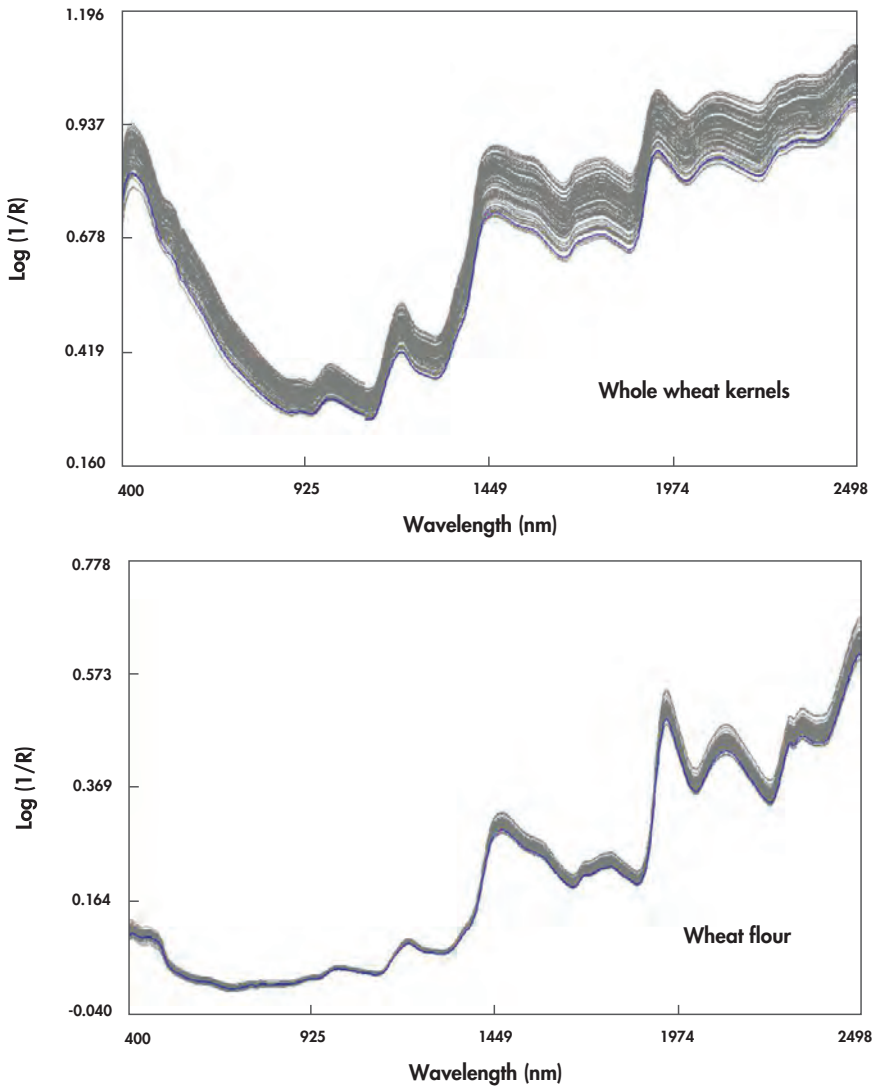


Figure 6.3 Spectra of whole kernels of Hard Red Spring wheat and flour (milled from the same wheat).

In transmittance mode some of the signal probably passes through the individual kernels or particles, as well as being reflected from their surfaces. The light energy can penetrate the kernels and even their cells. The same factors apply with regard to scatter. The sample thickness has to be optimised for analysis of different commodities. The reasoning for this is that there is a need for sufficient signal to reach the detector consistently. The scatter caused by large grains, such as maize, means that the sample has to be thicker so that enough signal diffusely reflected from the surfaces of the actual kernels gets through to the detector (rather than passing straight through gaps in the sample).

In the case of canola and sesame the seeds are very small. There is plenty of energy diffusely reflected from the individual seeds, but the scatter of the signal is so effective that the sample thickness has to be reduced so that sufficient signal can pass through to the detector (thicknesses of 3–6 mm have been reported). The sample thickness of wheat and barley at 18 mm, is intermediate. A rough calculation based on seed size would put about 6 canola seeds in a thickness of 6 mm, 7–8 wheat kernels in 18 mm and only about 5–6 kernels of maize in 30 mm. But the large surface area of the individual maize kernels would provide enough diffusely-reflected energy to provide for analysis by NIRS. Flour particles are even smaller than canola seeds, and a thickness of 2 mm is used for flour analysis by NIR transmittance.

Some whole-grain seeds, including soybeans and sunflower introduce another feature which may be related to some components of the seed coat that inhibit absorbance in some way. As a result, the optimum sample thickness for NIR transmittance analysis of whole soybeans is 30 mm, the same as for maize although the seed size is much smaller, and the shape quite different. In the case of whole soybeans and sunflower seeds, the characteristic oil bands in the 2300–2350 nm area are not apparent even though there may be respectively between 20% and over 50% oil in the sample. In both soybeans and sunflower seeds, the bands are present in the ground seeds. This shows that this apparent anomaly is a feature of the seed-coat and the difficulty of the measuring beam to penetrate the seed-coat to the germ, where the oil is located (Figure 6.4).

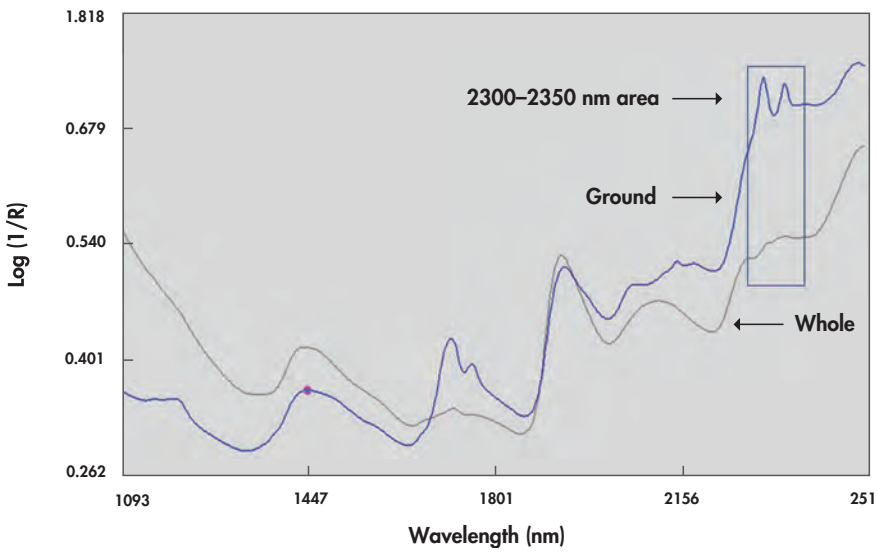


Figure 6.4 Log 1/R spectra of ground and whole sunflower seeds.

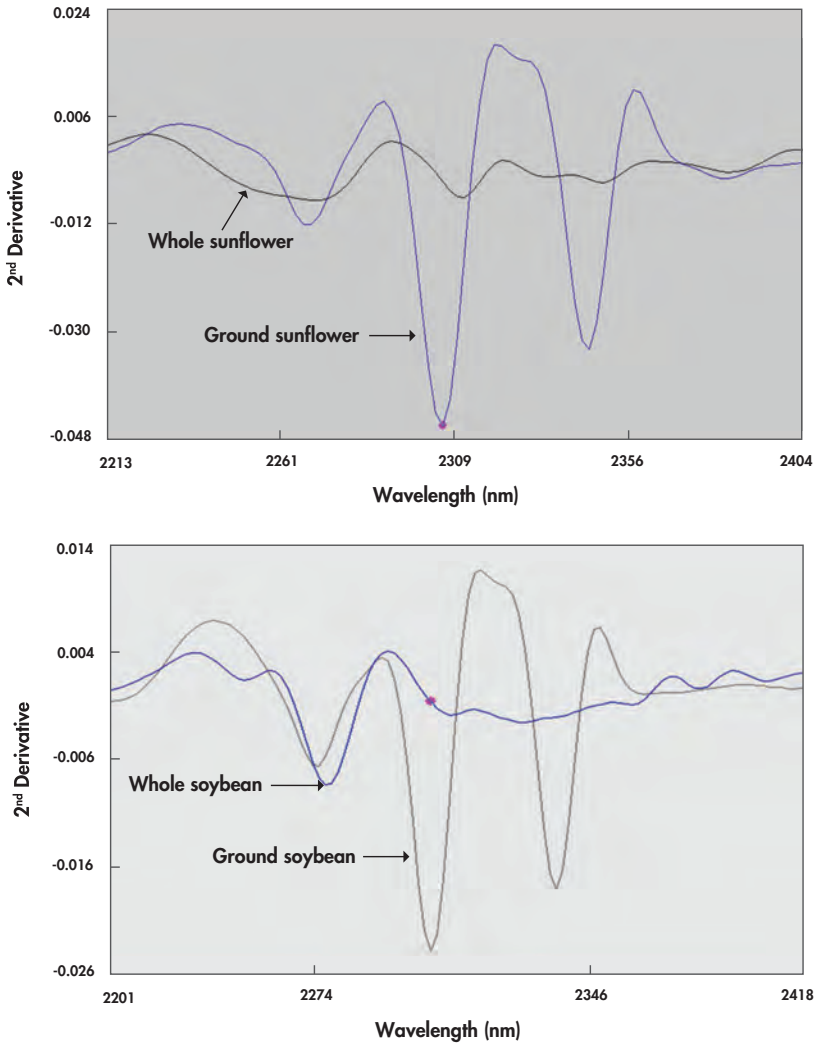


Figure 6.5 Second derivative (2 4 4 1) spectra of ground and whole sunflower seeds (top), as well as ground and whole soybean seeds.

Figure 6.5 shows 2nd derivative (2 4 4 1) spectra of ground and whole sunflower and soybean seeds in the 2300–2350 nm region. The designation 2 4 4 1 denotes second derivative, a gap and segment size of 4 wavelength points each, using WinISI software. The last digit of 1 indicates a segment of 1 which means that no second smoothing was applied. The prominent -CH_2 bands in the ground spectra are essentially absent from the spectra of the whole seeds.

6.2.3 Interactance

Interactance is another form of application of NIRS. Interactance measurements are made with a special fiber optics sensor containing two sets of fibers. A concentric bundle of fibres spaced a small distance from the centre bundle provides the

incoming radiation. The centre bundle of fibres collects the radiation transmitted into the sample and diffusely reflects back to the detector, which measures the interactance. The depth of penetration into the sample is a function of the distance between the two fibre bundles and the density of the material. This is illustrated in Figure 6.6. This arrangement provides for analysing the composition of tissue below the surface, and has to be optimised for different materials because the density of the material will affect the distance between the fibres. Interactance minimises the surface reflectance component if the sensor is placed in contact with the sample. Because of the small area of the sensing head, several scans should be taken and averaged for the most reliable results. The sensing head usually carries a sapphire window due to its resilience.

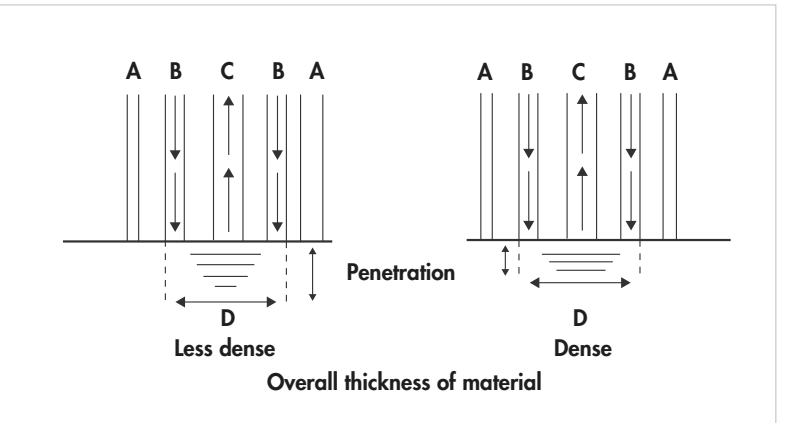


Figure 6.6 Illustrating the interactance principle. A = outer wall of cable; B = incoming radiation energy; C = diffusely-reflected energy; D = area of material scanned.

Optical fibres can be used in spectrophotometers to transmit radiation from one point to another. The absorption within the fibre material must be very low for this application, and in the case of NIR instruments this limits their use to shorter wavelengths or to shorter distances of up to about 5 m.

6.2.4 Instrument components

Apart from their physical size and appearance, the main differences among NIR instruments and whether they work in reflectance or transmittance mode, are the systems for wavelength selection, the detector, the presence or otherwise of an internal standard, the method of sample presentation and the system for instrument diagnostics. The instrument software is the software that instructs the instrument to record and store spectra, as well as translate spectral data into predicted reference data using the appropriate calibration. This software is similar

in principle for all types of instrument. Systems for wavelength selection include discrete narrow band-pass filters, acousto-optic tunable filters (AOTF), NIR light-emitting diodes (NIR LEDs), NIR diode lasers, moving-grating monochromators, fixed-grating monochromators with diode array detectors and interferometers.

Detectors

Detectors in most common use are the silicon detector in the wavelength range from 380–1100 nm, indium gallium arsenide (InGaAs) from around 900–1700 nm and lead sulphide (PbS) from 1100–2500 nm. Extended InGaAs detectors are available that extend their range to about 2400 nm. To cover the full wavelength range from 380–2500 nm, some instruments are equipped with both silicon and PbS or extended InGaAs detectors.

Sample presentation systems

Sample presentation systems include the method of sample access to the instrument and the system for irradiating (illuminating) the sample. Sample cells include static, moving or rotating cells of various shapes and sizes, with or without a glass or quartz cell cover. If the sample presentation system is designed so that the cell rotates on a turntable, the cell should rotate during the whole time of the scan. In this way the sample is stationary in the cell throughout the scanning and a large surface is presented to the detectors. Some instruments are designed so that the turntable stops between individual scans. If this is the case, the instrument will be programmed to process the spectral data at the instant that the turntable stops. This means that a smaller surface will be presented to the detector. Furthermore, the contents of the cell will not actually stop and will continue to move slightly for a few microseconds after the turntable stops. During this time, the instrument will be recording data on a moving sample. The combination of the smaller surface and the slight movement of the sample in the cell can have an adverse effect on the precision of the spectral data.

The sample can be radiated from either above or below. If the sample is viewed from above without a cell cover the surface will be more or less uneven, however, this can be compensated for by the rotation of the cell. If the sample is viewed from below, the surface will be the bottom of the cell, and will be planar (level). Provided the sample is viewed from below, open cups can be used with low-density fibrous materials such as wool and forages if a suitable weight is provided to compress the sample inside the cell. This is not practicable if the sample is viewed from above and viewing the sample from below gives more flexibility to the sample presentation system. If the sample cell can be used with a transparent

cover illumination from above is possible, but differences in the amount of sample loaded introduces another variable. The same principles with regard to path length apply when the sample is irradiated either from above or below. Some instruments are set up so that the sample is presented in a cell that moves vertically or horisontally, rather than rotating.

On-line instruments

For the purpose of this book, 'on-line' means that the instrument is installed to analyse in-stream, continuously during processes. These can also be called 'in-line'. This method calls for very rapid scanning which can be provided by diode array and AOTF instruments. 'At-line' means that the instrument is installed near the process, so that samples can be withdrawn and analysed whenever necessary. They may or may not require an external computer. Several sensing heads can be multiplexed to a single instrument and computer, e.g. for quality control within a processing plant.

Stand-alone instruments or at-line instruments

'Stand-alone' instruments are instruments that have been calibrated off-line. The calibrations are installed into the instrument, either manually, or if the instrument has a RS-232 or equivalent port, the calibration can be entered directly from an external computer. These instruments contain software to scan samples and compute results and need no peripherals.

Other at-line instruments are calibrated using off-line software that is installed into a computer. The instrument has software to record spectra and analyse future samples using calibration models. The external (generic) software uses spectral data recorded by the instrument using its data-storage options. The spectral data are transferred to the computer that carries the generic software and regressed against reference data to develop the calibration model. The model is set into the instrument that then uses it to analyse fresh 'unknown' samples.

Discrete filter instruments

The original discrete filter instruments contained up to 20 discrete narrow band-pass filters, each of which covered a wavelength range with maximum efficiency at a designated wavelength such as, e.g. a 1940 nm and a 2310 nm filter for moisture and oil analysis respectively. Narrow band-pass filters record the maximum absorbance at the designated single wavelength, but useable signals can be obtained down to the half peak-height (about ±5 nm) from that wavelength. No signals were recorded for wavelengths that occurred between filters. Some of

these used only 3 filters, but by tilting the filters through 90° a greater wavelength range was accessed than by ordinary discrete filter instruments. These instruments were programmed to record spectral data only at designated points, called 'pulse points' that corresponded to protein, moisture and oil during the rotation of the filter wheel. No signal was recorded from wavelengths that were not specified. The instruments used the log 1/R signal in the form of a precursor of the 1st derivative, and were all calibrated using MLR (multiple linear regression). The energy source of the first generation of whole-grain analysers came from NIR LEDs (near-infrared light-emitting diodes) and worked in transmittance mode. Subsequent generation of transmittance instruments used monochromators.

Grating monochromator

The introduction of the moving-grating monochromator in 1978 opened a new world of applications. Until then, NIRS technology had been limited to prediction of constituents such as moisture, protein, starch and oil in grains and seeds, as well as the same constituents plus ash in flour. The grating monochromator covered a larger and continuous wavelength range and allowed prediction of more complex constituents, including dietary fibre and individual amino acids and fatty acids together with some functionality parameters. It also made possible the use of PCA (principal component analysis) and PLS (partial least squares) regression for development of calibration models.

There are two forms of grating monochromator. The moving grating monochromator is used in most computerised spectrometers. The fixed-grating monochromator works with diodes and is the wavelength selection medium of most diode array instruments. Diode array instruments have no moving parts in the electro-optical system and are very fast in analysis, taking about 600 scans, and averaging them to give about 30 spectra per second. Many on-line applications use diode array instruments for continuous analysis of materials on moving belts or other systems. On-line instruments are usually calibrated with static samples. The calibrations are then tested and monitored using samples taken off the line, to verify that the calibrations are working properly.

Interferometer

Another type of instrument is the interferometer which can be used in the NIR or mid-IR range. The mid-IR range extends from about 3000 to 12,500 nm (3333 to 800 cm^{-1}). The terminology used to describe wavelength range in these instruments is reciprocal centimeters (cm^{-1}). In terms of reciprocal centimetres, the effective NIR range (from 800–2500 nm) extends from 12,500 to 4000 cm^{-1}.

The relationship is non-linear and wavenumbers can be converted to wavelengths using the formula,

$$x \text{ nm} = 10{,}000{,}000 / y \text{ cm}^{-1}$$

or wavelengths can be converted to wavenumbers using the formula,

$$y \text{ cm}^{-1} = 10{,}000{,}000 / x \text{ nm}$$

Appended to this chapter is Table 6.5, which translates reciprocal centimetres (cm^{-1}) into nanometres (nm) over the NIR range. The table has been organised to provide a difference of 2 to 4 nm per wavelength point. Spectral data of the interferometer are preprocessed using Fourier transform (FT), and the instruments are referred to as FT-NIR or FT-IR instruments depending upon the wavelength range. Sample presentation for these instruments used to limit their applicability to liquids, but recent advances have resulted in the introduction of a turntable/ rotating cell technique that has opened up their field of application to solid samples, including grains and pellets, forages, textiles and slurries as well as liquids. Wavelength precision of interferometers is superior to that of the more familiar dispersive ('normal') instruments.

Instrument noise

'Noise' is defined as any electronic signal that reaches the detector that is not directly related to the actual spectral data of the sample at absorption bands required for calibration and subsequent analysis. It is produced as a by-product of other activities in the system. Although NIR instruments are designed to reduce noise to the minimum, noise is always present to some extent in any NIR reflectance or transmittance system. Some instrument company brochures will state the signal-to-noise ratio of their instrument. The ratio of S/N is the ratio of the useful signal, in terms of the spectral data, to the electronic noise that is corrupting the signal at that wavelength point. The S/N ratio is greatest at the wavelength of the maximum radiation, and this is the figure reported by the instrument maker (usually around 1640 nm). It should be as high as possible. Typical S/N ratios for NIR instruments range from 25,000:1 to over 100,000:1.

6.3 Criteria for instrument selection

Successful application of NIRS depends to a certain extent on choosing the instrument most suitable for the application. Considerations in selecting an instrument are summarised in Table 6.1. These are not in order of importance – all of them are important.

Table 6.1 Factors to consider in selection of an NIR instrument

Factor	Explanation
Instrument purpose	Why is NIRS needed? What measurements are needed? In what material? What speed of analysis is required?
Instrument economics	Instrument price, in relation to intended application. What is the present cost of analysis, and how can an NIR instrument reduce costs and improve profits?
Instrument type	What is the mode of operation (e.g. monochromator, diode array, interferometer)?
Speed of testing required	On-line testing requires a very fast instrument, such as a diode array or AOTF instrument
Instrument size	How much bench space is required?
Internal reference standard	External reference standard is inconvenient to operator
Spectral range	Relates to type of analyses needed
Sample presentation system	Expected sample type and size; sample cell design; method of accessing sample to the cell, continuous, on-line testing
Instrument software	Dedicated or generic software for calibration model development
Calibration system	Transferability and efficiency of networking
Availability of factory calibrations	Factory calibrations are available for many factors
Operating manual	Often needs explanation to include more details
Simplicity in use	Are calibration models developed with instrument software? Are diagnostics adequate? Is it easy-to-use for non-technical people?
Durability	How long can an NIR instrument be expected to work? What is the life-span of energy source, detector(s) and sample access system? Ease of replacement of lamps.
Spectral quality	Are there noisy areas of spectra?
Instrument peripherals required	Keyboards and monitors use extra bench space.

Table 6.1 Factors to consider in selection of an NIR instrument (continued)

Factor	Explanation
Suitability for field operation and portability	Can the instrument be battery-operated? Is the instrument weather, dust and vibration resistant? Can the instrument be taken on-site to the field?
Instrument technical support	Company support for provision of calibrations, problem-solving and repair.
Instrument cost	Price of instrument relative to value of testing. This is more difficult to confirm if the instrument is to be used for analysis of materials such as waste products, soil, or for research.
Instrument networking	Important for clients with multiple instruments.
Diagnostics and maintenance	Simplicity in both of these criteria is an advantage

6.3.1 Purpose for the instrument

The first question to ask is 'why is an NIR instrument needed?' The question can be answered mainly by two things – the economics and the convenience. From the economic aspect the questions are 'by how much will NIR technology help to cut down operating costs and improve the efficiency of the operation?' and 'how can revenue be generated by using NIR technolgy?' The convenience questions are answered by 'what can be tested with NIR technology that would be very useful, but not possible without it?' and 'what is the benefit of getting the results immediately when they are most needed vs. getting the results in several hours or maybe even days'? Also, 'would the instrument be adaptable to continuous on-line testing?' and 'can the instrument be used in the field, i.e. away from the laboratory or other building?' The speed and low cost per test of an NIR instrument also offers the opportunity of testing raw materials, finished products, and in-process materials more frequently than was practicable with customary 'wet chemistry' analysis.

There is an increasing demand for information on composition, and functionality potential in industry, including modern grain handling and commodity processing to the extent that some form of analysis is becoming essential. Continuing quality and composition in a flour or feed mill or other type of processing plant must be assured. NIRS analysis can be very useful in determination of composition, both at the time of delivery of raw materials for pricing and during the manufacturing process. Constituents such as protein, oil and starch affect the value of grain, and instant on-site analysis by NIRS provides excellent potential to reduce costs of

monitoring and even generate revenue. Grain is stored for marketing, marketed directly or purchased as an ingredient for feed manufacture, and a price for a commodity such as grain has to be established. The moisture content of grains is an important criterion if grain is to be stored for periods of more than a few days.

In a food- or feed-processing plant, ingredients need to be tested to assure that they are suitable for use, and the progress of the processing needs to be monitored, including the composition of the final products. Materials such as manures that are stored in tanks or lagoons need to be tested continuously during pump-out and application to fields, because the manure in the storage facilities varies widely in composition. The same applies to sewage and waste water disposal.

Fresh fruits and vegetables need to be screened for grading before marketing. An NIR instrument can fill all of these needs (and many others) quickly and cheaply. Fresh forages used in the biogas industry need to be tested for moisture content, ideally during harvest, because payments are made on the basis of dry matter. Meats of all types, including fish, must be tested for quality, including parameters such as fat content, moisture loss during processing and marketing, and parasites.

The efficiency of soil analysis has always suffered from the costs of sampling and the subsequent analysis. NIRS analysis can reduce the analytical costs substantially, thereby reducing most of the costs of soil analysis to those involved with sampling. It enables testing of all samples taken from a field, rather than testing a composite prepared from all samples, which masks the variance in the field. Analysis of substances such as waste waters for biological oxygen demand (BOD) and other parameters is also handicapped by the costs of analysis and the long duration of conventional tests. The NIR technique is applicable to these materials and their associated tests.

Speed of analysis for the most part means whether the test is to be done on a stationary or moving sample. If the testing has to be done on-line, a diode array instrument is most suitable because of its speed. If the testing is to be done on a stationary sample, such as at grain receiving areas, feed or flour mill laboratories there are more choices of instrument. No matter how fast is the scanning, an NIRS test using a static at-line instrument takes about one to two minutes, because the sample has to be put into the instrument and taken out of it, and where necessary the sample cell has to be cleaned. A difference of a few seconds in actual scanning time has little effect on the overall time per test. Continuous on-line sampling eliminates the need for sampling, because all of the material is being analysed.

6.3.2 *Economics*

Having decided that an NIR instrument is required, the next thing is to work out the economics. Table 6.3 gives guidelines for the items to include in costing. Costs per test come down as the daily work-load increases. Table 6.3 is based on an expensive instrument, priced at US$90,000.00, and 1, 10, 20 and 50 tests per day. A survey of costs charged by commercial laboratories for reference testing revealed costs per test of US$24.00 for protein, US$13.50 for moisture, US$25.00 for fat/oil and US$25.00 for fibre for a total of US$87.50 per sample for proximate analysis (excluding ash and non-fibre carbohydrate material). The cost of testing varies, depending on the difficulty and expense of chemicals of the test. Some reference tests, such as metabolisable energy, cost US$800.00 or more per sample. NIRS tests all cost the same for every constituent. The costs of instruments and reference testing will vary (and usually increase). Table 6.3 is included to illustrate the difference between NIRS and reference testing. Total costs for an NIRS analysis of the same sample show that for an operation that carries out only 10 tests per day, the cost per test is a fraction of that of reference testing in a commercial laboratory. Remember that the cost per test includes all constituents and parameters tested (up to 12), because the NIR instrument tests them all simultaneously.

The costs reflected in Table 6.3 can be expected to increase over time, but the costs of NIRS testing will always be substantially lower than those of reference 'wet chemistry' testing, and the value in quality control of obtaining data continuously when needed is a further major benefit to any operation.

Table 6.3 Cost per test estimates for NIRS testing at different daily workloads

Number of tests per day	1	10	20	50
Instrument purchase price = US$90,000.00				
Instrument cost[1]	34.62	3.46	1.73	0.69
Calibrations (purchased)[2]	2.31	0.23	0.12	0.05
Calibrations (developed)[3]	21.88	2.19	1.09	0.44
Instrument set-up[4]	6.80	0.68	0.34	0.14
Labour[5]	0.68	0.68	0.68	0.68
Rental of laboratory space	1.53	0.15	0.08	0.03
Power	0.10	0.10	0.10	0.10
Total	**67.92**	**7.49**	**4.14**	**2.13**

[1]Values are in US$, amortised over 10 years; [2]Based on an average cost of a single calibration, amortised over 1 year; [3]Includes reference analysis on 200 samples; [4]Based on 20 min set-up and maintenance time per day; [5]Based on 2 minutes per test

Some instrument companies offer calibrations for a variety of commodities and constituents and functionality factors, such as metabolisable energy. Clients need to verify the accuracy of the calibrations by sending samples of their own commodities to a laboratory with which they are familiar. Biases may appear that have to be corrected since biases can become expensive. Operators can update their calibrations by adding their own samples with reference data. Note that calibrations developed in the laboratory are more expensive than factory calibrations.

To summarise the economics, NIR technology supplies the operation with very rapid on-the-spot testing with the same accuracy as reference testing, with equivalent or improved precision and at a fraction of the cost. An operation that has to obtain results from more than 20 tests each day is well-advised to use NIRS, and operations that need 30–40 or more tests per day really cannot afford not to use it. For an operation that handles commodities worth hundreds of thousands or even millions of dollars in a single year, the initial cost of a NIR instrument is a key investment in improvement of the operation.

On-site, on-line NIRS testing involves operations such as mounting an NIR instrument directly onto a combine or forage harvester and analysing the crop continuously during harvest, or continuous testing of liquids during pumping operations. In countries where large areas of grain are planted, mapping of fertility levels of very large fields increases the efficiency of fertilizer use. Use is already being made of protein-based fertility mapping in North America. Farmers who have been using the on-site NIR system claim savings of up to 20% in nitrogenous fertilizer costs. With the prices of nitrogenous fertilizers reaching, e.g. US$1000.00 per tonne (about US$0.98 per kg of nitrogen) it is easy to see the economic benefits of a 20% saving, when areas of 1000 ha or more have to be fertilized.

6.3.3 Instrument type

You do not need an axe to cut flowers. The type of instrument to be purchased will depend on the volume and complexity of the daily workload and on the precision needed. For example, for testing grain and flour for protein and moisture, simplicity of operation and speed of testing are very important. These can be met with stand-alone instruments. On the other hand, in a feed mill where up to 20 ingredients have to be compiled into a single feed mix and blended, a more powerful instrument, such as a monochromator-, diode array-, or interferometer-driven instrument is recommended. For an operation that is contemplating the use of NIRS for analytical work, it is important to think ahead to what the instrument

might be used for in the future, so that the appropriate instrument is selected even if the immediate outlay for the instrument appears to be higher than expected when the idea of using NIRS was initially presented to Management.

If very high-speed analysis is required, diode array and AOTF instruments are appropriate. These instruments are equivalent to other NIR instruments in precision or accuracy (in terms of SEP and bias), and the speed of analysis (only two to three seconds) makes them particularly suitable for some processing and for continuous and other on-line applications. There is increasing interest in FT interferometers. This technique called FT-NIR technology, is becoming increasingly applied to the analysis of agriculture and food materials. The instruments have excellent wavelength precision and could prove to be the NIR instruments of choice for the determination of some functionality factors, such as metabolisable energy.

For research work and prediction of functionality and other difficult parameters, a comprehensive instrument is recommended. The NIR technique is attractive to universities and other areas where basic research is carried out because a large amount of work can be completed in a short time, and scientific papers can be compiled in a few weeks. Wavelength ranges and mathematical pretreatments can be optimised and simulated. It is important to relate the results to eventual practical application under industrial and commercial conditions. This often involves design and testing of sample presentation systems, and the engineering aspect may become more important than the spectroscopy. When research reveals new areas where applications of NIRS appear to show potential, consideration must be given to the economics involved in the engineering and practical application. The 'bigger picture' includes consideration of the effects of using NIRS in place of the more labour-intensive conventional methods of analysis.

6.3.4 Speed of analysis required

Most benchtop instruments take about 1.5 to 2.0 minutes per test. No matter how quickly the actual scanning is done, the sample has to be placed in a sample cell, and the cell loaded into the instrument. After the scan is completed the cell has to be removed, emptied and the sample replaced into its original container. The sample cell usually had to be brushed or wiped clean before the next sample. The components of the loading and unloading operations usually take more time than the scanning, hence the 1.5 to 2.0 minutes. The difference in speed per testing comes when testing is to be carried out continuously on-line. No sample loading or unloading is necessary, nor are any samples needed because the whole population is being scanned continuously. But the speed of testing must be very fast so that

computation of composition is made on the basis of many scans. Diode array and AOTF instruments are most suitable for on-line work, but diode array instruments are generally less expensive and equally efficient. Advantages of continuous on-line analysis include elimination of the needs for sampling (because everything is being scanned) and bench space (because the instrument is either incorporated into the system by way of the production or pumping line).

6.3.5 Instrument size and portability

Instrument size can be important. In some areas bench space is often at a premium, such as grain elevators, feed mills, and other processing plants. NIR instruments are small – the largest is only about 60 cm long by about 35 cm high. No fume or dust exhaust or drainage is required, so installation costs are negligible. Some diode array instruments and some other instruments can be battery-driven, are fully portable, and can be taken into the field for testing materials, such as grapes, forages, soil on-site, and prospecting for different ores. For continuous operation in the field instrument, size becomes a factor in the insertion to operating equipment. Figure 6.7 shows a diode array instrument installed on a forage-harvester. The instrument was installed to measure moisture content during harvesting.

Figure 6.7 Diode array instrument installed on a forage harvester.

6.3.6 Internal reference standard

Most instruments are equipped with an internal reference. In the absence of such an internal reference the operator has to determine light and dark standards manually. These should be recorded at intervals of no more than 15 to 30 minutes. Manual recording of light and dark standards is an inconvenience to the operator, particularly at busy periods. The instruments make all measurements by comparison of the spectral data for each new sample with those of the most

recent standards, until the next standards are run. Minor changes in the spectral signal of the reference may occur between reference scans. Provided they are small and un-biased, it should not cause problems with the operation.

Ideally the reference standard should be run before every sample. For at-line instruments this can be done while the sample cell is being emptied, cleaned and re-loaded. The reference material in most older instruments was a ceramic tile, but in modern NIR instruments the internal reference most widely used is a durable plastic, called Spectralon®. Some instruments use an external standard that is exposed to the atmosphere. This is usually a ceramic tile. Because of dust build-up the surface must periodically be cleaned. Whether this is done by brushing or by means of a tissue, the action aggravates static electricity build-up, and a ceramic tile is less susceptible to static electricity than a plastic standard.

Transmittance instruments use air as a reference standard and a reference reading is taken automatically between each sample. On-line transmittance instruments can operate for long periods without taking new reference data.

6.3.7 Spectral range

The spectral range covered by commercially-available NIR instruments extends from 380–2500 nm. The spectral range can add substantially to the price of an instrument if more than one type of detector is required. Unless there is a need to measure colour components, such as chlorophyll, there is no need to spend the extra money on a wavelength area that will not be needed. Interferometers cover the range from 850–2500 nm (11,800–4,000 cm^{-1} expressed as wavenumbers). This wavelength range, or 1100–2500 nm, is adequate for most purposes. The wavelength range covered by the InGaAs detector (900–1700 nm) is becoming more widely exploited. This, coupled with their very practicable speed per test (about two seconds), makes diode array instruments very attractive.

The wavelength range scanned by individual NIR instruments depends on the purpose for which the instrument has been designed. For working in transmittance mode, the low wavelength range from 700–1100 nm is preferred. The main reason for this is that the absorbers in that region are all of low intensity, so that it is possible to use a thicker sample. The sample thickness must be optimised for different commodities. In the higher wavelengths, above 1100 nm the absorbers are stronger, and the sample thickness would need to be very thin (2 mm or less) for transmittance work. With such small diameter sample cells, it is difficult to fill them quickly, and also more difficult to clean them between samples.

Although some instruments that work in the low wavelength range operate over 400–1098 nm, the signals between 400 and about 650 nm tend to be noisy, and the most practicable range is between 650–1098 nm. For working with a coloured constituent such as chlorophyll, the instrument should be capable of scanning efficiently down to 674 nm (a prominent band for chlorophyll). For more complex applications, such as prediction of fibre components and energy, access to more wavelengths and more advanced mathematical pretreatments is helpful.

6.3.8 Sample presentation system

Some whole grain analysers operate over a relatively small wavelength range, can be delivered with factory calibrations, and have very simple sample access. The operator simply has to pour the sample into a hopper at the top of the instrument, the sample is scanned in about 40 seconds, the results are displayed and the sample falls into a drawer at the bottom of the instrument. This is ideal for elevator operators, who do not have time to load and clean a sample cell. Sample access is also simple in instruments that use an open sample cell revolving on a turntable and viewed from below. The lower surface is planar so the sample can be poured into the cell and emptied out after analysis. Sample thickness should ensure complete diffuse reflectance. This calls for a sample thickness of 3 cm or more, in the case of wheat and seeds of similar size, and at least 5 cm for maize, which has bigger kernels, or chopped forages, to ensure that light energy is diffusely reflected to the detectors, and is not wasted, by passing right through the sample.

Samples are presented to most instruments in a sample cell which usually has a window. Glass and quartz are two materials widely-used as sample cell windows. There are differences between these and among different individual glass or quartz windows. Figure 6.8 shows differences between quartz and glass windows and Figure 6.9 shows differences that can occur between individual quartz windows. Such differences can induce very large biases in the prediction of composition or functional parameters (see Chapter 7, Tables 7.3 and 7.4). Operators should check spare windows to ensure that they are compatible with the windows in use. Instrument companies should test deliveries of window materials before using them in sample cell assembly. Glass and quartz can both differ in composition, and the windows can also differ in thickness, both in average thickness, and in thickness in different areas of the window. Glass windows differ to a greater degree than quartz windows, both in composition and thickness.

Quartz windows are more susceptible to static electricity than glass windows. Static electricity is caused by transfer of electrons from one substance to another. It can cause samples to become orientated in direction, and samples of ground grains and forages can become arranged in certain positions on the surface of the cell window, which is what the instrument scans. These differences can cause biases. This is the reason why sample cells should be rotated through at least 180° when they are being scanned for analysis. This source of error is eliminated when sample cells automatically rotate. Oats are particularly prone to static electricity, and are arguably the most difficult of all cereals in sample preparation.

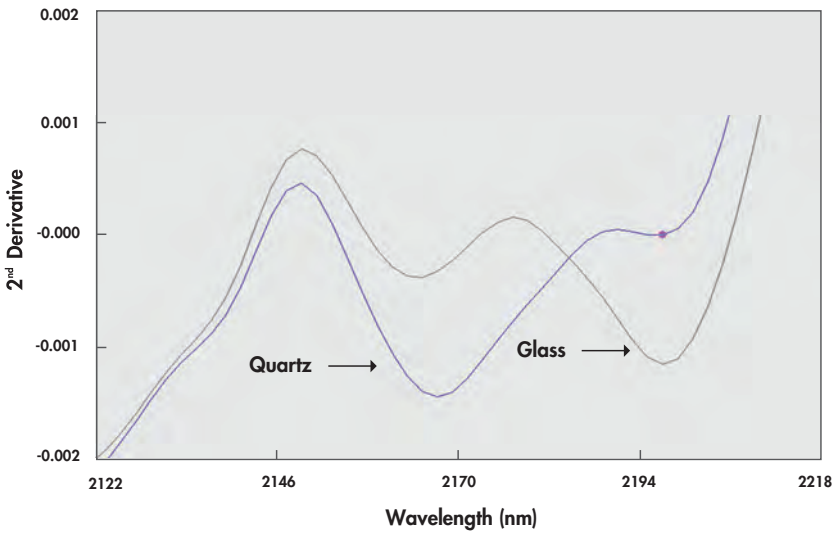


Figure 6.8 Differences in 2nd derivative (2 4 4 1) spectra of glass and quartz.

Whether or not the instrument uses a cell, every instrument has a 'window' of quartz or glass to protect the detectors from the atmosphere. These windows are subject to the same variability in composition.

The design of the cell has its own influence. Some sample cells are easier to use than others. The cell design also has a significant influence on the amount of sample that the instrument scans, and therefore on potential sampling error. The area of sample scanned also affects the precision of scanning. This is summarised in Table 6.4. Note that the precision of the spectral data for the rectangular cell was nearly three times better than that of the spinning cup.

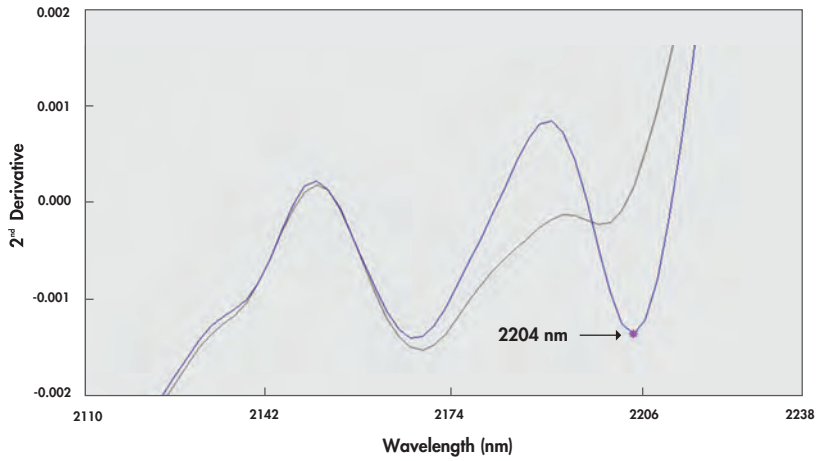


Figure 6.9 Differences at around 2204 nm between 2nd derivative (2 4 4 1) spectra of two quartz windows.

Table 6.4 Influence of sample cup size on sample size and instrument precision

Cell type	Area (cm^2)*	Number of kernels	Weight of grain (g)**	Mean log 1/R	SD***	CV (%)
Spinning cup	4.9	45	1.51	0.3018	0.00987	3.27
Rectangular	33.8	235	7.87	0.3027	0.00366	1.21

*Area of surface presented to instrument; ** Estimated weight; ***N = 12 scans

6.3.9 Instrument software

All NIR instruments have their own internal software that drives every function of the instrument in the same way that our own brain directs our hearts to keep on beating and our senses to keep functioning without our conscious control. The software instructs the instruments to take actions, such as revolving a turntable, recording and storing spectral signals and also responding to the commands of the external software that is used to record and translate the spectral data into analytical data.

Generic software is software that has been designed and prepared for many applications, of which only one is to translate NIRS spectral data into analytical data. For instruments that use generic software, the operator has to transpose the spectral data from the format in which they have been recorded into the format of the generic software, for calibration model development. The models are developed, evaluated and validated in the generic software and then re-entered into the instrument where the model efficiency is tested by analysis of completely

new samples. The Unscrambler®, GramsTM, MATLAB® or Pirouette® can be used to develop calibration models, using spectral data generated by the instruments with their internal software.

Dedicated software means that the software has been designed to enable the individual instrument to record spectral data and to employ calibration models to generate analytical data, the models having been developed by the same software chemometrics. Comprehensive dedicated instrument software has been developed by several instrument companies that market computerised spectrometers, most of which permit the development of calibration models. Examples of dedicated software are WinISI, Opus, and NIRCal. Some companies market their instruments factory-calibrated for prediction of frequently sought constituents or parameters, and operators can avoid the burdens of calibration. Software is discussed in more detail in Chapter 5.

6.3.10 Calibration systems and transferability among instruments

Calibration is a burden that NIRS technology has been forced to bear ever since its introduction. It is discussed in detail in Chapter 9 and is only briefly introduced here. The systems used for calibration development vary widely among instruments, but the principles are the same. Most stand-alone instruments are used in operations such as at grain delivery and purchase. These are all calibrated externally, using generic software, but can be delivered pre-calibrated for some constituents.

ANN (artificial neural networks) calibrations were first introduced about 25 years ago. This method requires a very large database, but ANN has the benefit of stabilising the calibrations and facilitating transfer of calibrations among instruments. Some grain operations maintain several hundred instruments all networked and use the same ANN calibrations. Instrument performance is monitored from the central laboratory of the company. The ANN calibrations originally developed at Tecator use thousands of spectral data collected from Infratec instrument users world-wide, under many operating conditions in terms of temperature and relative humidity. By using spectra from a very large number of instruments with accompanying reference data, the final calibrations compensate for the variance among instruments and laboratories, and the calibrations are very stable. One difficulty with the ANN approach is that of acquiring reliable reference data for constituents and parameters other than the main components of commodities, such as moisture, protein, oil and starch contents.

Transferability of calibrations means using the same calibration(s) on a network of instruments. Transferability of calibrations among instruments is very important in an operation that is using or proposing to use a series of instruments of the same type and model, because the output of all of the instruments can be monitored and adjusted from a single central desktop or laptop computer. To achieve this, ideally all of the instruments should produce identical spectra from a given sample and all of the instruments should yield the same results for the same samples. One successful method, is to standardise all of the instruments to a single Master instrument. A sealed cell containing test material is prepared and scanned and an equation developed and the cell and equation are used to standardise as many instruments as necessary. In this way, all of the instruments are then standardised to the master instrument, and can use the same calibrations.

Instrument companies can develop and expand their databases by arranging with clients to receive spectra with reference data. Organisations that operate more than one instrument can combine spectra and reference analysis to increase their databases for any constituents or parameters that they routinely measure. Merging spectra and samples with different reference analysis creates a form of 'sample noise'. While reference data from different laboratories may not exactly agree with one another, provided that there is no one-directional bias, this procedure will have the effect of stabilising calibrations. PLS regression, particularly the 'Local' version of PLS developed by WinISI, has also improved the reliability of calibrations.

It is rare to find two instruments that are exactly the same. Very small differences in gratings can cause significant differences in the output of even grating monochromators. Some manufacturers recommend that calibrations for a set of samples developed on up to five instruments are transferable to all other instruments of that type. Discrete filter and particularly NIR LED-based instruments are prone to even more variability, and these instruments are not preferred for networking.

6.3.11 Availability of factory calibrations

Some instrument companies offer factory-developed calibrations for commonly-sought constituents, such as protein and moisture contents. Such calibrations have been developed by experienced technical staff, using adequate numbers of samples. Some operations would prefer to use these calibrations, rather than undertake the development of calibrations within their operation. It is imperative that these operations verify the accuracy of the calibrations for their own requirements.

6.3.12 Instrument operating manual

Operating manuals supplied by the instrument companies are often lacking in key areas of day-to-day operation that can cause frustration to operators, and complicate efficient training. It is good practice for users to determine every step involved in every aspect of instrument operation, then prepare a detailed step-wise manual for their own training purposes. The operating manuals should include even the smallest details of operation, as well as how to apply corrections when an error has been made.

6.3.13 Simplicity in use

The main steps in day-to-day use of an NIR instrument are to check that the instrument is working properly before starting the important analysis of the day. This can be done by scanning a check sample or a check cell, and periodically using the diagnostics. These steps should be described in detail in an instrument operating manual. Instruments that employ a hopper-type sample presentation offer a big advantage in simplicity for the user. The sample can just be poured in, the analysis started by pressing the appropriate button, or even by simply touching the screen. The sample passes through the instrument into a tray when analysis is complete.

For instruments that use a sample cell, the method of filling the cell is a source of error, and the operating manual should include precise instructions for this all-important part of the test, particularly if the cell has to be cleaned between samples, e.g. for slurries and liquids, such as manure, high moisture samples, or very dusty samples. The easiest type of sample cell to fill is the open cup type, which is placed in a rotating turntable and scanned from below.

Simplicity in use also includes use and interpretation of the instrument diagnostics. These are important in assuring that the performance of the instrument optical system is functioning consistently. Some instruments contain a built-in system that tells the operator to run diagnostics. Some instrument companies offer clients a check cell option. The client submits a sample of the commodity or commodities for which they want a check cell, and the company prepares the check cells. When the check cells are received, they should be scanned 10 to 12 times to determine the mean and SD of the results. Thereafter the check cells should be scanned each day, ideally at different times, to expose it to different working conditions that may develop during the day. Provided the results are consistent the check cell approach can reduce the need to run full diagnostics to once or twice a week. If

the results of the check cell begin to show excessive variability the diagnostics should be run. A check cell can be prepared by any laboratory, provided that it is sealed, and protected from change over a reasonable period of time, such as 4 weeks.

Some instruments that scan from below, using a rotating turntable offer the option of using small cells, such as scintillation vials. For these instruments a vial can be used to prepare a check cell. The cell is filled to the top with a check sample of the test material, then the cap sealed in place with silicone grease to prevent changes in moisture content. The check cell is scanned first thing every morning to check that the instrument is working properly, and ideally scanned a few times during the day. The results should remain consistent for several weeks. If the results begin to change slightly the cell can be re-filled with the same check sample, and used over, because it is the bottom of the cell that is being scanned. If this system is used it is important to re-fill the same vial, to avoid possible differences between the composition of vials. Differences among vials will include differences in composition, and more importantly differences in the thickness of the bottoms of the vials. Variance introduced by these differences will automatically be incorporated into the calibration if a number of vials are used in the scanning for calibration development.

6.3.14 Instrument durability

Instrument durability is determined by observing how the instrument performs over prolonged use under normal fluctuations in daily workloads and operating conditions. Components that can affect the performance by aging include the energy source (usually a tungsten halogen lamp), the detectors and the sample access system. Modern NIR instruments work well for at least 10 years, and some of the older discrete filter instruments have provided good service for over 30 years. Maintenance is important to prolong the life of the instrument. This is usually limited to daily or at least weekly vacuum cleaning of cooling fans and careful attention to instrument diagnostics. It is useful to keep detailed, dated records of the output of check cells and diagnostics. These are valuable sources of information to servicing technicians if the instrument requires repair.

6.3.15 Spectral quality

Spectral quality can be studied by viewing the spectra throughout the wavelength range of the scan. Whenever working in an 'unknown' area, the spectra should be viewed to find if there are any noisy areas. This can best be determined by developing the 2nd derivative of the spectra. This will emphasise noisy areas.

Figure 6.10 shows 2nd derivative (2 4 4 1) spectra of some commodities over the 400–700 nm range. All of these spectra were taken from whole-grains or seeds. The spectra become smoother above 700 nm. Some instruments show very noisy spectra at both low and high wavelength extremes. Variation in colour adds to the factors of particle size and shape in inducing spectral noise.

The format in which instruments store spectral data influences the appearance of the spectra. Some instruments record spectra at 0.5 nm intervals, some at 2 nm intervals and others at 5–6 nm intervals. The latter spectra look smoother than spectra recorded at 0.5, and 2 nm intervals. These spectra are 'flattered' because their manner of storing spectra is essentially a form of smoothing. By smoothing the 0.5 nm and 2 nm spectra to the same degree as the 5–6 nm spectra, the spectra look just as smooth.

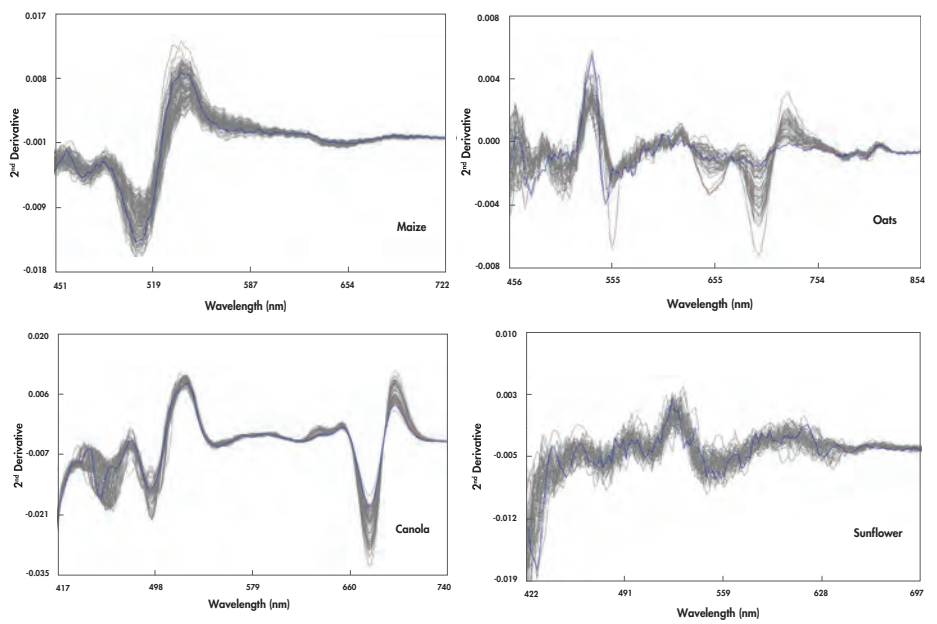


Figure 6.10 Noise in 2nd derivative (2 4 4 1) spectra in the low wavelength range.

Interferometers operate over the wavelength range from 850–2500 nm, but their spectral data are recorded as cm^{-1}. There is increasing interest in the use of FT-NIR instruments in agricultural applications. This comes partly from improvements in sample presentation. Interferometers, use different principles of operation. The degree of resolution of the interferometer can be specified by the operator, but as the degree of resolution is reduced the S/N ratio decreases, and resolution below 8 cm^{-1} is not recommended. Resolution of 16 cm^{-1} is adequate for most applications in the agricultural, feed, forage and food industries. Some FT-NIR software

includes an optimise option. The instruments scan the entire wavelength range. The software will select the best wavelength and mathematical pretreatment combinations for the sample sets presented. For this step the larger the database the better. The optimisation can be set up to use cross-validation or a test set. With a big data set (500 or more) the optimisation can take 30 minutes, or even longer, but the information gained is a valuable guideline.

6.3.16 Instrument peripherals

Earlier filter instruments and modern whole-grain analysers require no peripherals – a big advantage in operations where bench space may be limited. Stand-alone instruments are preferred for high volume operations, such as grain elevators and processing plants. Scanning monochromators, and similar instruments require peripherals such as an external computer, keyboard (in the case of desktop computers) and mouse, which increase the bench space required. Together with their associated cables, these peripherals are a nuisance in a high-volume operation. Provision can be made to accommodate the peripherals in the most efficient way so as not to impede the operation. For example, all cables can be located behind benching, and computers can be operated using a wireless mouse.

6.3.17 Suitability for field operation

In addition to the above specifications, for operation on-site in the field, e.g. for testing soils or forages, instruments have to be capable of being driven by battery or generator. Such instruments should be capable of being carried in a back-pack, including a laptop. Laptop computers have a limited operating time, but intermittent use can provide plenty of time for operating in the field. Several handheld instruments are available for field-operation. These instruments are designed for prediction of a small number of constituents, such as sugar or water contents. For these types of field-operated instruments, the instrument is presented to the sample, rather than the sample being presented to the instrument.

For sewage, or liquid manure analysis during application, the instrument should be capable of direct attachment to the pumping system or applicator. The system should include a valve for removal of samples for monitoring. The instrument should also be weather-proof, and tolerant of significant changes in ambient temperature, relative humidity, dust and vibration. Instruments with few or no moving parts are best adapted to these conditions. Diode-array instruments, with their high speed of analysis, are suitable for this type of work that is growing in application as more people become aware of the scope of NIRS analysis.

6.3.18 Instrument technical support

In selecting an NIR instrument, consideration should be given to the performance of the instrument company in providing follow-up advice and problem-solving after installation of instruments, in supplying and updating calibrations and in providing timely repairs. Instrument companies vary widely in the quality of after-sales service and in the expertise of service technicians. Most offer training in the basic operation of the specific instrument. One of the important areas that is generally lacking in the industry is thorough tutoring in the technology and how best to use it.

6.3.19 Instrument cost

It was mentioned earlier that the cost of the instruments should be considered in the context of the intended application and in terms of direct benefits such as cost-savings and producing a more dependable product. Another aspect of assessing the value of using NIRS is the consequence of continuing to not test, or not test with sufficient frequency due to the inconvenience, time and/or expense of testing. The purchase price of NIR instruments has often been a barrier to potential users of NIRS. Modern NIR instruments can be expected to function accurately for 10 years or more. Although the cost of the instrument has to be paid in its first year, the overall benefits make the initial purchase price an excellent investment.

6.3.20 Instrument networking

For operations that install more than a single NIR instrument, it is useful to be able to manage all of the instruments using a single calibration, or set of calibrations, for each constituent. The principle of networking instruments was first established a number of years ago and has since become widely used. Some networks consist of several hundred instruments, all of which are controlled from a single computer located in a central laboratory.

Networking and calibration transfer work best if the instruments all provide essentially the same spectral data from a given sample. One instrument is identified as the 'master' instrument, and all the others as 'slaves', or 'satellites'. The master instrument is calibrated to the reference methods. The satellite instruments are all calibrated from and to agree with the master instrument. Instrument accuracy and precision are monitored via the network using check samples that are either generated locally (the cheapest way) or sent out from the central control laboratory (the best way because only one laboratory is used for checking the reference data). The NIRS-predicted results for check sample analysis by the

satellite instruments are compared to the NIRS-predicted results from the master instrument for the same samples, rather than to the reference laboratory results.

Changes in satellite instrument performance that require adjustment to slope or bias are all carried out on the basis of the results from the master instrument, and they are all effected via the network. In practice, slope changes are very rarely required and are not recommended. Monochromator-driven instruments are preferred for networking. Differences among monochromators and interferometers are generally smaller than differences among discrete filters, CCDs or diode arrays. NIT monochromator-driven instruments use air as the internal reference and are the most suitable for large-scale networking. Networking is discussed in more detail in Chapter 9 (see section 9.2).

6.3.21 Instrument maintenance and diagnostics

Simplicity on both instrument maintenance and diagnostics is an advantage to day-to day operation. NIR instruments require regular maintenance which usually means removal of dust, preferably by vacuum. A record should be kept of when the instrument first began to be used, so that an estimate can be made of the probable life of the lamp (light or energy source). Most lamps have a life expectancy of about 2000–2400 hours, and some modern instruments use lamps with 10,000 hours expectancy. It is important to turn the lamp off after use, unless the instrument is in constant use, either on-line or because of shift-work. When a lamp has to be changed, allow it to warm up for at least one hour then scan the check cell and two or three check samples a few times in order to determine whether the new lamp has introduced a bias. Also note the date of the installation of the new lamp in order to estimate its probable life.

The instrument should never be turned off. Instruments take about 1–2 hours to stabilise when they are newly turned on and analysis may not be reliable during that period. Moreover, diagnostics may show that the instrument is not performing up to standard if it is turned off every night and turned on again in the morning. Software is available to turn instruments on for a set time before the workday begins to allow the instrument to warm up. By reducing its use this prolongs the life of the lamp by up to 50%.

6.4 Last words on choosing an NIR instrument

Choice of an NIR instrument depends on the use to which it is to be put. Some instruments have been designed for rapid, simple analysis of commodities such as grain. Most of these are NIT instruments. Others are designed for more complicated

analysis, such as prediction of functionality. Still other instruments are bought just for research into the field of NIR technology.

The wavelength accuracy and spectral resolution of mid-IR and FT-NIR and FT-IR instruments are regarded as being superior to those of conventional NIR instruments. The mid-IR manufacturers have overcome initial problems associated with sample presentation, and there is increasing interest in application of mid-IR in several fields of agriculture. Nevertheless, for a technique to supersede another it has to prove itself to be superior, rather than equivalent, in crucial areas of application (including costs). The consensus among long-time users of NIRS is that for some applications, especially in the feed industry, FT-IR and FT-NIR instruments will continue to gain ground, but for most applications in agriculture the 'classical' NIR technique will continue to be the method of choice for the next few years mainly because of its flexibility in applications, as well as the wealth of information that is available on calibration transferability and networking and instrument durability.

The golden rules for successful NIR instrument application:

1. Select the instrument appropriate to your applications – usually by discussion with the instrument company.
2. Develop a sample presentation system that can be used consistently with more than one operator.
3. Identify or develop a reliable laboratory for reference analysis and determine precision of all reference testing.
4. Establish check samples for every commodity that will be tested, including whole and processed materials – bulk check samples should be very thoroughly blended, and stored in airtight containers, ideally refrigerated but not frozen – subsamples can be taken from time to time for daily use.
5. Make sure that the instrument spectral data are precise using 'check' samples and a sealed check cell – keep dated records of check cell and check sample results.
6. Make sure that calibration models are accurate and precise, also using 'check' sample.
7. Make sure that factory calibration models perform accurately on your material.
8. Develop a reliable monitoring system.
9. Make sure that all operators are thoroughly trained in the operation of the instrument, and that they understand the statistics needed for monitoring of accuracy and precision.
10. Carry out diagnostics regularly and keep records of diagnostic performance;

and prepare a comprehensive operating manual containing all steps of the operation.

Most on-line applications are customised, particularly with respect to sample presentation, and would-be users are advised to discuss their specific needs with instrument companies that offer this type of equipment. Calibrations are developed using samples withdrawn from the on-line system. Individual instrument manufacturers will advise how manual samples for calibration and monitoring should be taken.

Table 6.5 Conversion of wavenumber (cm^{-1}) to wavelength (nm)

Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)
12500	800	11280	887	10560	947
12460	803	11240	890	10540	949
12420	805	11200	893	10520	951
12380	808	11180	894	10500	952
12340	810	11140	898	10480	954
12300	813	11100	901	10460	956
12260	816	11060	904	10440	958
12220	818	11020	907	10420	960
12080	828	11000	909	10400	962
12140	824	10980	911	10380	963
12100	826	10960	912	10360	965
12060	829	10940	914	10340	967
12020	832	10920	916	10320	969
11980	835	10900	917	10300	971
11940	838	10880	919	10280	973
11880	842	10860	921	10260	975
11840	845	10840	923	10240	977
11800	847	10820	924	10220	978
11760	850	10800	926	10200	980
11720	853	10780	928	10180	982
11680	856	10760	929	10160	984
11640	859	10740	931	10140	986
11600	862	10720	933	10120	988
11560	865	10700	935	10100	990
11520	868	10680	936	10080	992
11480	871	10660	938	10060	994
11440	874	10640	940	10040	996
11400	877	10620	942	10020	998
11360	880	10600	943	10000	1000
11320	883	10580	945	9980	1002

Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)
9960	1004	9100	1099	8240	1214
9940	1006	9080	1101	8220	1217
9920	1008	9060	1104	8200	1220
9900	1010	9040	1106	8180	1222
9880	1012	9020	1109	8160	1225
9860	1014	9000	1111	8140	1229
9840	1016	8980	1114	8120	1232
9820	1018	8960	1116	8100	1235
9800	1020	8940	1119	8080	1238
9780	1022	8920	1121	8065	1240
9760	1025	8900	1124	8050	1242
9740	1027	8880	1126	8035	1245
9720	1029	8860	1129	8020	1247
9700	1031	8840	1131	8005	1249
9680	1033	8820	1134	7990	1252
9660	1035	8800	1136	7975	1254
9640	1037	8780	1139	7960	1256
9620	1040	8760	1142	7945	1259
9600	1042	8740	1144	7930	1261
9580	1044	8720	1147	7915	1263
9560	1046	8700	1149	7900	1266
9540	1048	8680	1152	7885	1268
9520	1050	8660	1155	7870	1271
9500	1053	8640	1157	7855	1273
9480	1055	8620	1160	7840	1276
9460	1057	8600	1163	7825	1278
9440	1059	8580	1166	7810	1280
9420	1062	8560	1168	7795	1283
9400	1064	8540	1171	7780	1285
9380	1066	8520	1174	7765	1288
9360	1068	8500	1176	7750	1290
9340	1071	8480	1179	7735	1293
9320	1073	8460	1182	7720	1295
9300	1075	8440	1185	7705	1298
9280	1078	8420	1188	7690	1300
9260	1080	8400	1190	7675	1303
9240	1082	8380	1193	7660	1305
9220	1085	8360	1196	7645	1308
9200	1087	8340	1199	7630	1311
9180	1089	8320	1202	7615	1313
9160	1092	8300	1205	7600	1316
9140	1094	8280	1208	7585	1318
9120	1096	8260	1211	7570	1321

Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)
7555	1324	6910	1447	6334	1579
7540	1326	6895	1450	6322	1582
7525	1329	6880	1453	6310	1585
7510	1332	6865	1457	6298	1588
7495	1334	6850	1460	6286	1591
7480	1337	6835	1463	6274	1594
7465	1340	6820	1466	6262	1597
7450	1342	6805	1470	6250	1600
7435	1345	6790	1473	6238	1603
7420	1348	6775	1476	6226	1606
7405	1350	6760	1479	6214	1609
7390	1353	6745	1483	6202	1612
7375	1356	6730	1486	6190	1616
7360	1359	6715	1489	6178	1619
7345	1361	6700	1493	6166	1622
7330	1364	6685	1496	6154	1625
7315	1367	6670	1499	6142	1628
7300	1370	6655	1503	6130	1631
7285	1373	6640	1506	6118	1635
7270	1376	6625	1509	6106	1638
7255	1378	6610	1513	6094	1641
7240	1381	6598	1516	6082	1644
7225	1384	6586	1518	6070	1647
7210	1387	6574	1521	6058	1651
7195	1390	6562	1524	6046	1654
7180	1393	6550	1527	6034	1657
7165	1396	6538	1530	6022	1661
7150	1399	6526	1532	6012	1663
7135	1402	6514	1535	6002	1666
7120	1404	6502	1538	5992	1669
7105	1407	6490	1541	5982	1672
7090	1410	6478	1544	5972	1674
7075	1413	6466	1547	5962	1677
7060	1416	6454	1549	5952	1680
7045	1419	6442	1552	5942	1683
7030	1422	6430	1555	5932	1686
7015	1426	6418	1558	5922	1689
7000	1429	6406	1561	5912	1691
6985	1432	6394	1564	5902	1694
6970	1435	6382	1567	5892	1697
6955	1438	6370	1570	5882	1700
6940	1441	6358	1573	5872	1703
6925	1444	6346	1576	5862	1706

Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)
5852	1709	5444	1837	5100	1961
5842	1712	5436	1840	5092	1964
5832	1715	5428	1842	5084	1967
5822	1718	5420	1845	5076	1970
5812	1721	5412	1848	5068	1973
5802	1724	5404	1850	5060	1976
5792	1727	5396	1853	5052	1979
5782	1730	5388	1856	5044	1983
5772	1733	5380	1859	5036	1986
5762	1736	5372	1862	5028	1989
5752	1739	5364	1864	5020	1992
5742	1742	5356	1867	5012	1995
5732	1745	5348	1870	5004	1998
5722	1748	5340	1873	4996	2002
5712	1751	5332	1875	4988	2005
5702	1754	5324	1878	4980	2008
5692	1757	5316	1881	4972	2011
5682	1760	5308	1884	4964	2015
5672	1763	5300	1887	4956	2018
5662	1766	5292	1890	4948	2021
5652	1769	5284	1893	4940	2024
5642	1772	5276	1895	4932	2028
5632	1776	5268	1898	4924	2031
5622	1779	5260	1901	4916	2034
5612	1782	5252	1904	4908	2037
5602	1785	5244	1907	4900	2041
5592	1788	5236	1910	4892	2044
5582	1791	5228	1913	4884	2048
5572	1795	5220	1916	4876	2051
5562	1798	5212	1919	4868	2054
5552	1801	5204	1922	4860	2058
5542	1804	5196	1925	4852	2061
5532	1808	5188	1928	4844	2064
5524	1810	5180	1931	4836	2068
5516	1813	5172	1933	4828	2071
5508	1816	5164	1936	4820	2075
5500	1818	5156	1939	4812	2078
5492	1821	5148	1943	4804	2082
5484	1823	5140	1946	4796	2085
5476	1826	5132	1949	4788	2089
5468	1829	5124	1952	4780	2092
5460	1832	5116	1955	4772	2096
5452	1834	5108	1958	4764	2099

Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)	Wavenumber (cm^{-1})	Wavelength (nm)
4756	2103	4428	2258	4162	2403
4748	2106	4420	2262	4156	2406
4740	2110	4412	2267	4150	2410
4732	2113	4404	2271	4144	2413
4724	2117	4396	2275	4138	2417
4716	2120	4388	2279	4132	2420
4708	2124	4380	2283	4126	2424
4700	2128	4372	2287	4120	2427
4692	2131	4364	2291	4114	2431
4684	2135	4356	2296	4108	2434
4676	2139	4348	2300	4102	2438
4668	2142	4342	2303	4096	2441
4660	2146	4336	2306	4090	2445
4652	2150	4330	2309	4084	2449
4644	2153	4324	2313	4078	2452
4636	2157	4318	2316	4072	2456
4628	2161	4312	2319	4066	2459
4620	2165	4306	2322	4060	2463
4612	2168	4300	2326	4054	2467
4604	2172	4294	2329	4048	2470
4596	2176	4288	2332	4042	2474
4588	2180	4282	2335	4036	2478
4580	2183	4276	2339	4030	2481
4572	2187	4270	2342	4024	2485
4564	2191	4264	2345	4018	2489
4556	2195	4258	2349	4012	2493
4548	2199	4252	2352	4006	2496
4540	2203	4246	2355	4000	2500
4532	2207	4240	2358		
4524	2210	4234	2362		
4516	2214	4228	2365		
4508	2218	4222	2369		
4500	2222	4216	2372		
4492	2226	4210	2375		
4484	2230	4204	2379		
4476	2234	4198	2382		
4468	2238	4192	2385		
4460	2242	4186	2389		
4452	2246	4180	2392		
4444	2250	4174	2396		
4436	2254	4168	2399		

07

Sampling, sample preparation and sample presentation

Sampling and sample preparation are two aspects of analytical work by NIRS or reference methods where research has seemingly lagged behind other developments in the technology. The main reason for this is that the requirements of sampling in industry have been recognised for many years, and the research essential to effective and efficient sampling has been accomplished accordingly. In most industrial operations involving large volumes of bulk materials, sampling is automated, and the precision and reliability of the systems have been scrupulously established.

There are a number of things that the average NIRS user needs to know about samples, sampling, sample preparation and sample presentation to the instrument. These include:

- What is a sample?
- What is the significance of a 'truly representative' sample?
- How should a sample be taken?
- What affects the accuracy of sampling?
- How many samples are needed for a calibration?
- How should samples be prepared for analysis?

- How should samples be stored?
- How is the sample presented to the instrument?

The spectra are the heart of NIRS application to any type of material, and factors that can affect the spectra have to be included in the assembly of samples. In the case of grains, factors such as variety, growing location and season can affect spectra even of samples with the same chemical composition as determined by reference methods. Sample assembly for calibration development should identify the most important variables, and assemble samples accordingly. Many of these factors are identified in Chapter 8. Before beginning a calibration exercise, it is essential to establish the sources of variance, and to identify sources of samples that represent the most important variance factors. Although this has nothing to do with the actual sampling, it is inherent to the success of sample assembly for development and evaluation of NIRS calibration models.

7.1 What is a sample?

Statisticians and analytical chemists look upon a sample in different ways. To a statistician a sample is a portion of a population that is intended to represent accurately the entire population in its original appearance, and physical, chemical and physicochemical form. To the analytical chemist, a sample is simply an envelope or other container full of material that has to be analysed.

The term 'population' is used to describe the total amount of whatever has to be tested, whether it is all of the kernels in a truckload of grain, all of the bales in a 50-hectare field of hay, the full amount of liquid in a lagoon of manure, or any amount that is too big for testing in its entirety. Ideally, the whole population should be tested to determine its composition or functionality. This is very rarely practicable with solid materials, except with small populations such as the seeds from a single plant. Because NIRS is non-destructive, the seeds can be planted after testing. NIRS has opened small doors, as well as big ones. Continuous on-line testing does permit analysis of total populations of liquids, slurries, and of solid materials in a moving stream. No samples need to be taken because the whole population is being tested so that sampling error is essentially eliminated.

Some confusion may arise about the use of 'N' or 'n' when discussing statistics. The value 'n' refers to the number of samples taken from a population, e.g. the number of samples taken to represent a cargo of 30,000 tonnes (metric tonnes) of grain or the number of bales that have been sampled to represent a field of 50 hectares. The value 'N' refers to the total population. If the sample size is 10 kg

of grain and a sample is taken for every 1000 tonnes, this means that 30 samples are assumed to represent the total cargo and n = 30, whereas N (the number of theoretical samples in the population) = 30,000,000 (kg) ÷ 10 = 3,000,000 an unrealistic number. Similarly in a 50 hectare field, that has yielded 4 tonnes of hay per hectare, and a sample is taken to represent the total yield of hay in the field, the total number of samples (N) will be (200 x 1000) ÷ 2 = 100,000 also an unrealistic number. For practicality all NIRS work 'n' is used more frequently than 'N'.

If the precision of the *sampling* is to be assessed when the samples are analysed by NIRS or reference methods, it is essential to know the precision of the *testing* methods. Here the CV (see Chapter 4, section 4.1.3) is useful. If the CV of the SEP (standard error of prediction) is high, based on NIRS or reference analysis, but the CVs of both of the SETs (Chapter 4, section 4.1.9) are low, it means that the sampling error is high. If this is the case, the reasons should be determined and corrected. Sampling error is part of the total error, and it is impossible to run an efficient analytical system if the sampling error is high.

Under laboratory conditions the actual sampling is generally not the responsibility of the analytical chemist or technician, whose duties are to analyse the samples that arrive in the laboratory and report the results. If the reproducibility of the test methods is reliable, yet the results prove unsatisfactory to the client, the sampling technique may require investigation. In some locations, such as at grain delivery points, where samples of farm trucks may have to be taken manually for testing for moisture, protein, starch or oil contents, the duties of the person that does the test may include that of taking the sample. It is important to establish a reliable sampling procedure at such locations. Many modern elevators have pneumatic probes for sampling farm trucks. At older elevators samples may be taken manually at the rear of the truck during unloading. This is acceptable provided that the hand-sampling is continued until the truck has completely unloaded.

7.2 The truly representative sample

The truly representative sample should provide the material for the most accurate method of determination of the composition and functionality of the entire population. To save expense and time in testing, fully representative samples are usually thoroughly blended composites of sub-samples that have been taken

Hand-sampling a farm truck at a grain elevator.

during operations such as unloading grain from railcars, storage of grain in silos, collecting straw bales from fields, pump-out of manure lagoons, sampling large fields for soil analysis, and other operations where the populations are too large and the number of sub-samples too numerous to test individually.

The danger with these truly representative composite samples of large bulks is that they mask the variance that occurs within the sample. For example, if a composite sample is taken during storage of farmers' deliveries of freshly-harvested grain into a big silo, the composite may show that the moisture content is 12.9%, a safe value for storage. But the individual deliveries may have ranged from 11 to 16% in moisture content. The delivery of, e.g. 12 tonnes of grain at 16% moisture into a silo could cause heating, mould development and spoilage within the silo, that would not become apparent until the grain was later moved out of the silo for marketing. The heated grain would then become dispersed throughout the grain as it moved out of the silo (Figure 7.1).

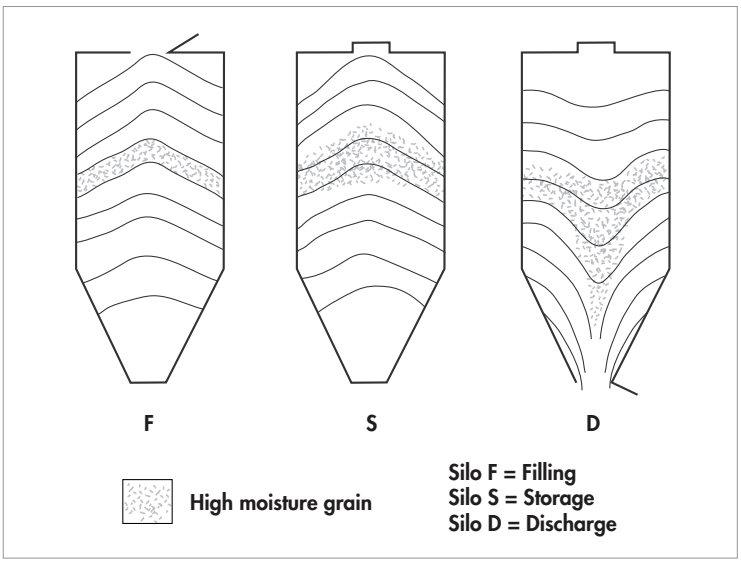


Figure 7.1 Spread and dispersal of a single delivery of high moisture grain to a silo.

When a composite sample is going to be prepared from a series of deliveries, sub-samples of cargo loading, or field samples, the variance or the SD (standard deviation) among the samples and the range in analytical data should be determined to answer any complaints. To obtain this important information, each individual sub-sample that goes to make up the composite should be tested before compositing. This was impracticable with reference testing, but it can be easily and cheaply achieved using NIRS by scanning every sub-sample as it is taken.

7.3 Sampling

The success of both NIRS and reference testing depends heavily on the integrity of the sample. On-line NIRS systems offer an advantage. The population is continuously scanned and spectra are recorded at designated intervals for NIRS analysis. No samples are needed for day-to-day analysis of liquids and slurries because all of the material is continually analysed. Nevertheless, samples have to be withdrawn for reference testing for calibration and performance monitoring. On-line systems are most efficient for application of NIRS where the depth of the sample stream is thin enough to preclude stratification, and monitoring the surface of the stream is sufficient to provide accurate analysis of the entire stream. These systems are less successful in grain streams because the streams are deep enough to cause stratification. Vibration of conveyor belts and other conveyors cause lighter components of the grain, such as weed seeds and chaff, to migrate to the surface and can influence the results. The sampling has to be carried out in such a way that the entire depth of the belt or conveyor is being sampled to avoid the error of stratification. The determination of such foreign material forms an important segment of the pricing so that accurate sampling for determination of the foreign material is essential.

Continuous on-line sampling can be achieved by means of sensors located above the stream. Alternatively, the instrument can be directly interfaced with the stream of material, solid or liquid, by positioning it so that the stream of material is passing directly past the sensor. This is best achieved by installing the instrument at an angle in ductwork, or at an angle slightly below 90° on the bottom half of circular ducting. Continuous in-stream NIRS testing of liquids, such as milk or liquid manure, is effective because the material is continuously moving and cannot settle out or stratify.

The two main sampling or sample withdrawal systems are manual and automatic. Manual sampling of grain can be carried out by probes or other specialised handheld sampling devices, or simply by diverting a stream of grain falling from a truck into a pail by hand. This method is effective, provided the diverting process is continued during unloading of the entire load and the sample is thoroughly blended before the NIRS analysis. Truckloads can also be sampled manually or automatically by means of hollow probes that are dipped through the volume of grain and used to withdraw grain at all levels of the bulk. Trucks and even railcars of the box-car type can be sampled in this way. In the case of box-cars or gondola cars, at least six probes should be taken to cover the full surface and depth of the bulk of grain. Open trucks can also be sampled in this way. Some receiving points

such as elevators and mills have pneumatic probes that draw the sample and deliver it directly to the testing area. Again, at least six probes should be taken. The six sub-samples should then be thoroughly blended before analysis. This can best be done by using a blending device such as a Boerner sample divider.

In large terminal elevators and most flour and feed mills, bulk grain is sampled by automatic samplers that continuously withdraw sub-samples from the main streams of grain, flour, feed mixes and other types of products. The sub-samples are sent to a pneumatic system for transport to the elevator laboratory. This method of sampling usually removes too much sample for practicable analysis. The streams of sampled materials are consequently passed through a stream-splitter that allows a certain percentage of the stream to be transferred to the laboratory – redirecting the remainder to the original conveyor.

The Boerner sample divider.

The two main types of samplers for grains are the diverter-type and the Woodside samplers. The Woodside sampler uses chains that move continuously through the stream of grain. Each chain carries small cups for removal of sub-samples. Some grains, such as peas, are very hard and difficult to sample accurately with a Woodside-type of sampler, because hard, round seeds tend to 'jump' out of the sampling cups. Diverter samplers move across the stream, usually at the end of belts or conveyors where the grain is being discharged into a hopper and divert sub-samples into the pneumatic transfer system.

Diverter samples are more efficient than Woodside-type samplers and are becoming the sampling method of choice. Still other processing plants, such as paper mills, monitor their streams directly with NIRS fibre-optic sensors installed over each stream. This calls for careful optimisation of the positioning and standardisation of all sensors (there are usually more than a single stream in a big plant), so that there are no biases among the sensors. Samples still need to be removed for determination of foreign material, which cannot be measured accurately or precisely with NIRS. Flour mills can either install NIR instruments on selected flour streams and monitor them continuously for protein, moisture and ash contents, or withdraw samples automatically into the laboratory. There is a very wide diversity of sampling systems custom-designed to serve specific purposes in industrial processing plants.

Moving belts can be sampled by hand samplers such as the Ellis cup, and open grain streams can be sampled by pelican samplers. These items of equipment are available from companies such as Seedburo. Sampling of bags of grain, pellets, flour or any other type of bagged material can be carried out by probes that penetrate the bags.

Sampling of soil in a field is usually carried out by means of an automatic auger or a manually operated auger. Soil sampling is complicated by the size of the fields, the topography and the need to take samples from at least two depths of the soil. The top 10–15 cm will contain roots, which should be removed before analysis of the soil itself, unless the soil is to be sampled for organic matter, including carbon. To obtain a reliable estimate of the composition and condition of the soil of a field of 100 hectares is a formidable task. Sampling of fields of big plants, such as whole maize (corn) plants is also difficult and operations that work with this type of material have to develop and evaluate their own sampling methods.

The Vertis soil sampler.

Forages are usually baled in the field before storage or use. Bales can be sampled by manual or mechanical probes that penetrate the bale and withdraw samples from throughout the bale. The location of bales in the field can be documented to allow an estimate to be made of the variability of forage all over the field. Wool and cotton bales are sampled in the same way.

The integrity of the sampling also depends on the efficiency with which the material, to be sampled, is actually representative of the population. If the material is stratified, the sampling must penetrate all of the strata. This becomes difficult, e.g. in the sampling of a lagoon of hog manure, where the surface may have total solids content of only 0.5–1%, whereas the bottom layers will be much higher in total solids and nutrients – so that the lagoon or storage tank must be continuously agitated during pump-out. Layering of liquids changes refractive indices. Temperature has an effect by changing the viscosity of slurries. All of these factors must be incorporated into samples assembled for calibration models for NIRS analysis of these types of materials.

7.4 What affects the accuracy of sampling?

The accuracy of sampling is affected mainly by five factors – the type of material, the inherent heterogeneity of the material, the location of the sampling system and the method and frequency of withdrawal. The type of material mainly concerns its physical characteristics. This includes whether it is solid, granular, liquid, slurry, powdered, fibrous, such as hay or any other forage, wool or cotton, fresh or dried, and other physical factors such as texture. It also includes the composition, particularly in terms of moisture and fat contents. All of these factors affect the way in which the sample can be taken and stored.

The location of the sampling systems will be decided by the operator and by the operation. Grain, flour, feeds and similar materials can be sampled from belts or spouting. Forages, wool and cotton and similar materials can be sampled from bales. Soils, manures, and composts can be sampled at the site by suitable augurs and probes. Frequency of sampling and sample size will be determined by the operator. Sample size depends on the heterogeneity of the material, and in general the higher the degree of heterogeneity, the bigger should be the sample, and the more frequently should samples be taken.

The **first step** in establishing a sampling system is to know the tolerance expected (sometimes called the 'degree of uncertainty'). This will affect the number of samples that must be taken. The stricter the tolerance, the higher is the degree of sampling needed. There is no need to be stricter than the operation calls for. For example, if the tolerance required in analysis of wheat for protein content is ±0.20, there is no need to strive for an SEP of 0.15%. This would be more demanding of staff, and more expensive. The **second step** is to determine the precision of the method of analysis that will be used to determine the heterogeneity, because without that figure it is impossible to determine the true heterogeneity. The **third step** is to determine the heterogeneity of the material. This can only be done by experiment. A **fourth step** is to determine the economics of the frequency of sampling and testing, because the costs of the system have to be feasible. Determination of the heterogeneity of a blended bulk can be carried out rapidly and effectively by NIRS with no need for a calibration. This is discussed in Chapter 11 (see section 11.11).

A useful system is to take frequent samples, retain and analyse each of the samples individually, and make a composite sample using a constant amount, or volume, as each of the individual sub-samples is added to it. Each of the progressive composites is analysed by the reference method, or by the NIRS method, assuming that you have a reliable calibration. It is quicker, cheaper and just as accurate.

Then calculate the mean of the individual samples (the weighted mean if the sub-samples differ in weight or volume). It should agree with the result obtained on the final composite sample. The SD of the results of the individual sub-samples shows the variance, and heterogeneity of the material. 'Normal' values for the CV for grain protein content should be between 1–2%. The CV for samples of soil, manures, forages and more complicated materials can be expected to be higher, but should be determined and should ideally be less than 5%. To determine the frequency of sampling, the number of individual samples used to calculate the mean is gradually reduced. The minimum number is reached when the mean does not change. If the mean of all of the individual samples fails to agree with the final composite, the frequency of sampling will need to be increased. When the frequency of sampling has been determined the entire exercise should ideally be repeated to verify the reliability.

The guidelines for sampling, presented in sections 7.1 to 7.4, apply particularly to solid materials. Liquids and slurries call for different methods of sampling. Water is the main constituent of most of these materials. Water contains no constituents except total solids, so the application of any analytical method, including NIRS, is for the determination of the amount and composition of the solid material present.

7.5 Sample preparation

In these days of direct NIRS testing of intact commodities such as whole grain, the errors introduced by grinding and some other aspects of sample preparation, are largely eliminated. But sample preparation is as important as ever for the reference analysis that is to be used for monitoring instrument performance, and for development and extension of calibrations. There are several steps in sample preparation, some of which are also applicable to whole-grain NIRS testing. These include:

1. Careful identification and labeling of samples at the time of sampling or receipt of the samples;
2. thorough mixing/blending of the sample before any testing is attempted;
3. testing for moisture content as received;
4. moisture reduction, as necessary;
5. accurate sub-sampling, if the sample is too big for processing;
6. removal of foreign material, unless the commodity is to be evaluated for pricing or functionality as received;
7. size reduction, usually by grinding or chopping, to the point where the material can be subjected to reference testing;

8. blending the sample after size reduction because grinding usually induces stratification;
9. cleaning of the grinder between samples where necessary;
10. testing for moisture content after sample preparation;
11. storage of samples after preparation and before NIRS or reference analysis, to avoid changes in composition;
12. labelling of the stored samples; and
13. refrigeration, as necessary, to protect from changes in composition in storage.

Steps 1, 2, 5, 6, and 12 are equally applicable to whole grain testing. Accurate documentation is the first priority. Any testing done on a sample with the wrong identification is a waste of time and money, and can have serious consequences. When assembling samples for developing calibration models, it is important to record the source and date (including the year), if the samples are to be preserved for more than one year. Thorough blending of the sample is also essential before any testing is done, particularly if the sample is a composite sample.

High moisture materials are the most difficult to handle. High moisture grain is difficult to grind, e.g. for reference analysis, and impossible to grind without variable and uncertain moisture loss. Furthermore, if grain is to be scanned using an NIR instrument after grinding, grinding high moisture grain changes the particle characteristics. This can change both slope and bias of the predicted results significantly. Grain that is high in moisture (above 12–13%) should be tested for moisture content by an approved moisture metre that has been calibrated to a standard method (such as Approved Method No. 44-15.01 of the AACCI), or ideally by a two-stage oven method before grinding, so that an accurate result can be obtained for moisture content. Using this method, a sample of the grain is weighed and air-dried to a lower moisture level to determine the stage A moisture level. The partly dried sample is ground and re-tested by air-oven to determine the stage B moisture content. The total moisture content is calculated by a formula that compensates for stage A (AACC, 2000). High moisture forages and manures should also be tested for moisture content (total solids in the case of liquid manures) before further processing.

For reference analysis of fresh forages, silages, vegetables, fruits, meat, fish, soils and manures, the sample preparation steps will differ. All have to be dried and processed before analysis. These materials are more difficult to prepare than high moisture grains. One system is to determine the moisture content and scan them the way they are at the time of receipt. Then sub-sample, dry and grind them

for analysis and determine the moisture content of the material at the time of reference analysis. The calibration model is developed using the scans on the fresh material, and the reference data have to be converted to the original moisture content. Fresh materials, such as forages, cannot be ground and have to be reduced to a size that can be dried and ground (for reference analysis) or scanned (for NIRS analysis). For large materials, such as whole grass or maize plants, a technique has to be developed within the laboratory and the spectral precision evaluated for reproducibility before scanning (see Chapter 9, section 9.1.2).

Whether or not results have to be reported on a constant moisture basis, such as moisture-free (dry) basis, it is important to test the moisture content before and after sample preparation. This minimises the risk of inducing significant and variable biases in the reported results. Much confusion in relating NIRS to reference data can be avoided by reporting all results on a constant moisture basis. Conversely, for preparation of samples such as feed mixes or computation of total dietary intake by cattle or sheep, where a large proportion of the intake may be as fresh feed, the composition with respect to nutrients (such as protein, fibre and energy) should be reported as a function of the actual composition of the feed ration. Fresh manures present a unique problem in that they should be scanned as received, which is an unpleasant operation in the case of solid manures. Drying such material for reference analysis also has to be carried out with care, ideally by low temperature or freeze-drying, to avoid losses of steam-volatile ammonia and other constituents.

In the following example, the actual protein content as consumed gives the meaningful information about the intake of protein by the animal. The composition of concentrate fed to the animal should be formulated to supplement the optimum diet of the animal, based on the intake of the fresh forage (with protein content of 5.7%).

Example: Fresh maize silage

Protein (reference)	= 21.4%
Moisture (at time of reference testing)	= 6.8%
Moisture as received	= 73.4% (26.6% dry matter)
Actual protein as eaten	= 6.1%

Careful blending of the sample before sample preparation is particularly important if a sub-sample has to be withdrawn, which is usually necessary. The method of sub-sampling is also critical, particularly with forages such as maize and other

large materials. It is preferable to take at least two sub-samples of these materials and test both. Size reduction usually means grinding. In the case of soil and similar materials, air-drying or at least low temperature forced-air drying, is necessary before size reduction can be carried out effectively. Air-dried soil can be ground in a hammer mill, or preferably by a specialised soil grinder (something like a motorised pestle and mortar) that does not exert such a vigorous effect on the physicochemical characteristics of the soil. Soil and commodities such as oilseeds are usually ground to pass through a 2.0 mm screen. Grains, including soybeans, are usually ground to pass through a 1.0 mm screen.

The Cyclone, Perten KT-3100 and Retsch centrifugal grinders are widely used in grain sample preparation. Of these, the Retsch grinder has proved to be the most suitable for working with oilseeds. Forages can be ground in any of the Wiley series of mills, or in a Christy-Norris 8-inch hammer mill. The Wiley No. 4 grinder is a large mill, and is easy to clean. The Christy-Norris mill is very robust, and can be used to grind seeds such as faba beans and palm kernels, which are extremely hard. This is a direct-drive mill, and noisy, so ear-protection equipment should be provided for operators.

Grinders vary and if more than one grinder is used, operators should check that spectral characteristics are not affected by differences in grinders. Figure 7.2 shows spectra of the same wheat sample ground on two Cyclone grinders. Differences of this order would cause significant biases, and would possibly affect slopes.

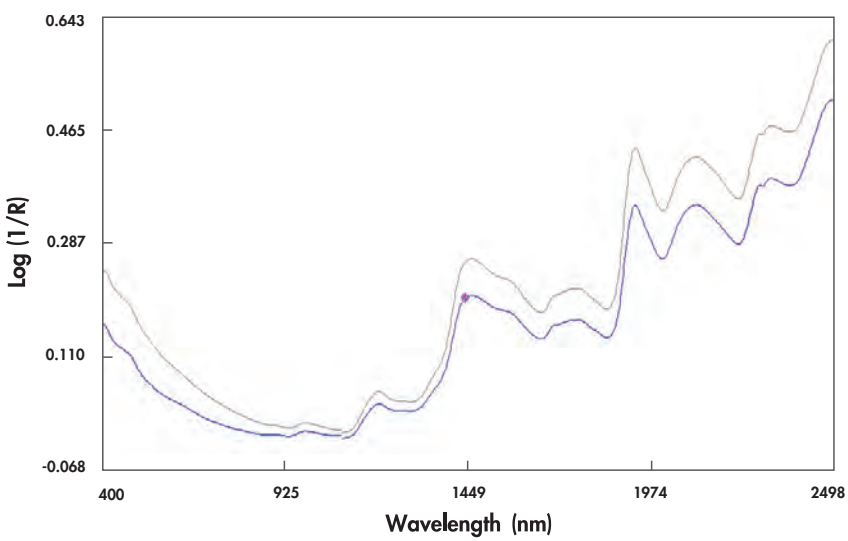


Figure 7.2 Spectra of the same wheat sample, ground on two Cyclone grinders both fitted with 1.0 mm screens.

Cleaning of grain before analysis is an essential part of sample preparation. Foreign material, such as weed seeds, seed pods and stones can interfere with NIRS spectroscopic measurements and should be removed before analysis unless the on-going samples will always be analysed on arrival with the foreign material included. Because the amount of foreign material affects the value and price of the commodity, it is important that the blending preserves the distribution of foreign material as it was in the original material as received. Large-scale cleaning of grains and seeds is carried out in grain-handling and processing plants, and detailed descriptions of the equipment and methods involved are outside the scope of this book.

Blending of ground material is another essential to successful analysis. The different strata of ground grains have different composition. For example, after grinding, the protein content of the finest particles can be 3–4% (absolute) higher than the mean protein content obtained after thorough blending. In preparation of dried forages, care has to be taken to preserve the ratio of leaf to stem during sub-sampling and sample preparation for reference analysis. Leaves are much higher in protein and lower in fibre contents than stems. A big advantage of NIRS testing is that the sample size that can be scanned by the instrument is usually much larger than the sample that can be subjected to reference analysis. For example, in testing grain for protein content by reference Kjeldahl or combustion analysis, the sample size is usually 250–1000 mg, whereas some NIR instruments can scan more than 100 g.

Cleaning of grinders between samples avoids sample-to-sample contamination. Some grinders, such as the Cyclone grinder and some hammer mills, are self-cleaning when grinding cereal grains and pulses. But when grains such as maize and oats are to be ground, where the oil content can be as high as 8%, most grinders require cleaning between samples. Oats are particularly prone to static electricity, and are arguably the most difficult of all cereals in sample preparation. Static electricity is caused when two surfaces come into contact with each other. Electrons can be transferred. Most of the time the electric charges remain in the materials, but if a material carrying a static electricity charge comes into contact with another surface, which has lower resistance to electrical conductivity, charges will be transferred. Sometimes, when the air is very dry, a metal spatula dipped into a sample of ground forages or oats will come out looking like a toothbrush. Here the Udy grinder suffers some disadvantage. Its plastic grinding chamber intensifies the presence of static electricity.

Grinders that are not self-cleaning have to be cleaned between samples using brushes and a vacuum. Air-streams are not recommended because they increase the atmospheric dust and can cause sample-to-sample contamination. Up to over 2% of samples of grain may remain in a grinder after grinding. These fine particles are of different composition from the blended ground material, and should be completely removed by brushing, and added to the sample before blending.

Precautions that are needed to avoid changes in composition are particularly important in sample preparation for reference analysis. The main reason for this is that operations such as moisture reduction and grinding are most likely to cause changes in composition with respect to volatile constituents. Low molecular weight fatty acids, and some of the esters that contribute toward flavor, are volatile in steam and are lost during removal of moisture by oven-drying at 100°C or lower. Precautions should be taken either to avoid such losses, or to determine these components separately. They can contribute an important amount of the digestible energy of some feeds. The Karl Fischer method for determination of moisture is preferred for some commodities. One problem with this method is that it depends on complete extraction of the water from the commodity. Extraction efficiency can be affected by particle size and texture.

7.6 Sample storage

Storage of samples is an important aspect of sampling. Samples require protection from change between the sampling and sample preparation stages and analysis, and possible re-analysis, by NIRS or reference methods. Many commodities can be safely stored in plastic bags. Under a low-power microscope a plastic bag looks like a chicken-wire fence. The 'gaps' allow water molecules to escape. 'Ziploc' bags can protect grain from changes in moisture content for several years, provided the bags are also stored in tight-lidded plastic pails, and at low temperature. Ziploc bags should not be stacked too tightly in boxes or pails because they may burst open. It is a good idea to insert sample ID inside plastic bags in case the labeling wears off from the outside of bags. Rigid plastic containers are available in many shapes and sizes. Tight-fitting lids impede changes in moisture content. Ground materials that are waiting for testing by reference or NIRS methods should be stored in containers that prevent changes in moisture content, such as plastic containers with tight-fitting lids, or in metal cans protected by taping the lids. Failure to protect the contents by taping will result in moisture loss. This is illustrated in Table 7.1.

Table 7.1 Moisture losses from ground wheat stored in metal cans with and without taping the lids

	Initial moisture content (%)					
	14.3		12.6		10.4	
	Moisture loss (%)		Moisture loss (%)		Moisture loss (%)	
Days from initial moisture test	**Taped**	**Not taped**	**Taped**	**Not taped**	**Taped**	**Not taped**
1	0	0.1	0	0.1	0	0
2	0	0.9	0	0.4	0	0.1
4	0	3.3	0	1.4	0	0.3
6	0	4.0	0	2.1	0	0.3
21	0.4	–	0.2	–	0.1	–

Table 7.2 Safe' moisture levels for grain storage at different environmental relative humidities and temperatures

	60% Environmental relative humidity		70% Environmental relative humidity	
Grain	20°C	30°C	20°C	30°C
	Safe moisture levels (%)			
Wheat	14.5	13.5	13.0	12.5
Barley	15.0	14.5	14.0	13.5
Oats	14.5	14.0	13.0	12.5
Maize	14.5	13.5	13.0	12.5
Rice	13.5	13.0	12.5	12.0
Sorghum	14.5	13.5	13.0	12.5
Millet	16.5	16.0	15.5	15.0
Soybeans	12.0	11.5	10.0	9.0
Dry beans	15.5	15.0	14.5	14.0
Sunflower	8.0	7.5	7.0	6.5
Flax	9.0	8.5	8.0	7.5
Canola	8.5	8.0	7.5	7.0

Storage of check samples is particularly important. The results of testing check samples cannot monitor reproducibility if the sample itself is changing. Check samples should be stored in heavy-duty plastic bags and refrigerated. Sufficient

sample should be brought out at each time to last for 2 to 3 days, and should be allowed to equilibrate to room conditions before scanning by NIRS. For prolonged storage (one week or more in warm climates) samples should be stored at 4 to 5°C. Table 7.2 gives the maximum moisture levels at which grains can be safely stored for up to 5 to 6 days at environmental relative humidity levels (ERH) of 70% and 60%.

Bulk density of samples has to be taken into consideration when samples are to be stored. Samples of 500–1000g of materials, such as soil or grain, have fairly small volume, and many samples can be stored in an area of a cubic metre or so. Samples of forages and similar light materials should also be of the same size, but require much more storage space.

7.7 Sample presentation to the instrument

7.7.1 Sample presentation for reflectance

The method for sample presentation to the instrument is crucial with both whole and ground materials. In reflectance mode, many instruments present samples of both whole and ground material in some form of sample cell. The design, size and composition of sample cells, is an essential aspect of instrument design. This is discussed in Chapter 6 (see section 6.2.8). Cell size has a significant influence on the amount of sample that the instrument scans and, therefore, on potential sampling error (see Chapter 6, Table 6.4).

Ground materials that are presented to the instrument in a sample cell are viewed by the instrument through a window, which is made either of glass or quartz. Both glass and quartz differ in composition, but glass windows vary more in thickness than quartz windows. Changes in composition affect the spectra of any materials scanned, and changes from sample cell to sample cell influence NIRS analysis if the windows differ. Changes in window thickness affect the effective distance between the sample surface and the detector. Even a small change of a few microns can change this distance enough to affect the results, and can change the path-length, which is the distance that the light energy travels inside the sample, before it is diffusely reflected to the detectors. Differences in the sample surface presented to the instrument can be caused by differences in window thickness. The errors can be detected by rotating a cell through 180°. Instruments that rotate the sample cell eliminate this error.

Tables 7.3 and 7.4 give examples of changes in bias and other statistics induced in wheat flour by differences in glass and quartz windows. Errors caused by

differences in the composition of sample-cell windows are not limited to glass. Although the biases in the quartz series appear to be lower than those of the glass series, they are very significant. Such errors also affect precision of NIRS analysis, as determined by analysis of check samples.

Table 7.3 Influence of glass window material on prediction of protein content (%) of flour where calibrations developed with an individual window were used to predict a sample presented in a cell with the same window and with a different window

	Calibrated with glass window #1 – Predicted with #1			Calibrated with glass window #1 – Predict with #5		
Wavelength range (nm)	**SEP (%)**	**Bias**	**RPD**	**SEP (%)**	**Bias**	**RPD**
408–1092	0.236	-0.005	8.8	0.297	5.036	7.0
1108–2492	0.097	0.007	21.5	0.107	2.126	19.5
408–2492	0.128	-0.003	16.7	0.130	-0.025	16.0
	Calibrated with glass window #5 – Predicted with #5			Calibrated with glass window #5 – Predicted with #1		
408–1092	0.347	-0.046	6.0	0.328	0.979	6.4
1108–2492	0.104	0.004	20.1	0.117	4.125	17.8
408–2492	0.143	-0.014	14.1	0.156	-0.231	13.4

Although a great deal of NIRS analysis is carried out on whole materials, NIRS is widely used in the analysis of commodities, such as flour and feed mixes, most of which are scanned through windows. It is important for any user who carries out analyses of any type that requires a quartz or glass window to verify cell-to-cell variance.

Because many instruments present whole materials to the instrument in some sort of sample cell, instrument companies should also determine the extent of variance in deliveries of glass or quartz for fabrication of sample cells or viewing windows, e.g. in setting up on-line systems where the material being analysed is scanned through a window.

The detectors of all instruments are protected from the atmosphere by a window of some sort, and these materials (special glass or quartz) can also vary. Because of this, instruments wherein the sample is scanned in an open cell without a window are not exempt from this source of error. The open cell approach also carries its own error sources due to changes in the surface during scanning.

Static electricity is a source of error with instruments that use a sample cell for analysis of flours or meals. When the sample is added to the cell the static electricity can cause the sample particles to become dispersed in random patterns at the surface of the window, quartz windows are more susceptible than glass windows to this phenomenon, and some commodities, such as ground oats and some ground forages are also more prone to static electricity than others. The changes in the surface presented to the instrument can cause biases. Rotating the sample cell through 180° can compensate for this source of error, provided that the duplicate results are averaged for reporting. Instruments that automatically rotate the sample also avoid this source of error.

Table 7.4 Influence of quartz window material on prediction of protein content (%) of flour where calibrations developed with an individual window were used to predict a sample presented in a cell with the same window and with a different window

	Calibrated with quartz window #1 – Predicted with #1			Calibrated with quartz window #1 – Predicted with #5		
Wavelength range (nm)	**SEP**	**Bias**	**RPD**	**SEP**	**Bias**	**RPD**
408-1092	0.343	-0.048	6.1	0.488	0.009	4.3
1108–2492	0.104	0.007	19.1	0.109	-1.061	18.0
408–2492	0.129	-0.020	16.2	0.120	0.723	13.4
	Calibrated with quartz window #5 – Predicted with #5			Calibrated with quartz window #5 – Predicted with #1		
408–1092	0.281	-0.029	8.0	0.283	0.021	7.4
1108–2492	0.107	-0.015	19.5	0.123	0.983	17.0
408–2492	0.137	0.001	15.2	0.158	1.074	13.2

The amounts of ground materials presented for scanning are smaller than those of intact materials, and are usually presented in covered sample cells. The layer of ground sample should be thick enough to prevent the instrument from scanning right through the sample and picking up signal from the back of the cell. The radiation from the illumination source will penetrate to about 2 mm of a finely-ground or powdered material before it becomes completely scattered. When analysing straws or forages, the cell must be filled sufficiently densely so that the instrument cannot scan through the sample. If this does happen, anomalies may appear in the spectra. Radiation penetrates whole grains to depths of a centimetre (or more in the case of large-grained materials, such as maize), and penetrates packed straw and similar materials to depths of several mm.

7.7.2 *Sample presentation for transmittance*

The most popular whole grain instruments employed in wheat and barley analysis work in transmittance mode. This type of instrument uses an incremental flow-through system for sample presentation. The number of increments is specified by the operator during calibration development.

In transmittance mode, sample thickness is even more important than in reflectance mode. The reason for this is discussed in Chapter 6 (see section 6.2.2). The optimum path length (thickness) is determined during calibration, and varies from about 2 to 3 mm for flours to 3 to 6 mm for small seeds such as canola or poppy seeds, and up to 30 mm for maize and soybean. The sample is viewed through a window of about 1.0 cm^2 so that the total volume of grain scanned at each increment (sub-sampling) ranges from about 300–3000 mm^3. In the case of wheat, for which a path length of 18 mm is recommended, the amount of wheat per increment is about 1.55 g, so that an analysis covering 10 increments will scan about 16 g of grain (even if the actual sample size is as much as 1 kg). Some instruments use a viewing chamber, the walls of which are not square, and the increments of grain are proportionately larger.

For constituents or parameters that are evenly distributed among individual kernels, as in the case of grains, scanning samples in increments of as small as 1.6–2.0 g is acceptable. The SD of protein content among individual kernels in a sample of wheat is about 0.6%, while for moisture content the SD is only about 0.4%. The respective CV values are about 9.5% and 3.2%. This means that most of the kernels in the sample have about the same protein and moisture contents. Scanning 10 increments provides very precise results. The same applies to kernel texture in ground wheat, where the CV is about 1.5%. If the constituent or parameter is more variable among kernels, the SD of the increments will increase correspondingly, and the results will be affected. This has been found to be the case with parameters such as Falling Number (FN) in wheat, an indication of sprout damage, or 'scab', i.e. *Fusarium* head blight, where only about 2–4% of the kernels may be affected. For such applications NIRS is only suitable for rough screening.

Another form of scanning is 'transflectance'. This method is used for the NIRS analysis of liquids. The sample is either loaded into a cell with windows at both surfaces, or it may flow through a transparent tube. By either method a highly reflective material or reflector is placed behind the cell (or tube). The sample is scanned through the cell. Energy is diffusely reflected from the reflector back through the cell and reaches the detector with the information that it has gained from the sample. When transflectance is used with a liquid that is likely to change

its composition to the extent that it becomes opaque the energy will not reach the reflector, and the energy then becomes diffusely reflected from the sample. This is illustrated in Figure 7.3. As the liquid contains progressively more particles, no energy passes through to the reflector.

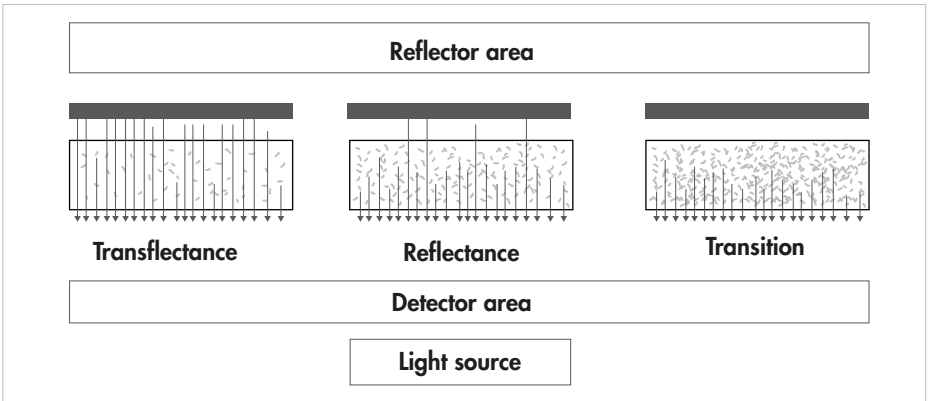


Figure 7.3 Illustrating transition from transflectance to total reflectance (transition).

Liquid manures and similar materials can be analysed continuously using a flow-through system. The material is viewed through a window, usually of sapphire. The system should also be fitted with apparatus for collecting a sample, for monitoring. The spectral signals are passed to the detector either directly, or via fibre-optics. The NIRS testing mode can change from a diffuse transflectance to a diffuse reflectance measurement, as the composition of the material changes, so the system should be fitted with a reflector, and the calibration models developed to accommodate both transflectance and reflectance modes. If the material is contained in piping or tubing of 10–20 cm, there is no need for such an external reflective material because the sample thickness is sufficient to prevent the signal from passing completely through.

7.7.3 Sample presentation on a turntable

Some instruments employ an open sample cell set onto a revolving turntable. This allows a larger amount of sample to be scanned. The sample cell can be made from metal, plastic or glass. The sample is scanned using a light source that irradiates an area of about 2.0–2.5 cm in diameter. For this type of instrument, if the sample is scanned in an open cell from above because the sample is rotating, uniformity of the surface is not really essential to reproducible scanning. The instrument software will accommodate unevenness of the surface as a variable. The problem with scanning from above lies in analysis of materials such as forages or wool, where the surface is very uneven, and possibly 'fluffy'. A uniform surface is

assured in instruments that use a transparent sample cup, illuminated from below. Forages, wool and similar materials can be compressed with a suitable weight, to assure an even surface. This approach, while assuring an even surface for scanning, is subject to changes in the material from which the window of the cell is made, usually special glass or sapphire. Instruments that use the circular open cell/turntable system of sample presentation also avoid some of the problems associated with loading high moisture samples.

The turntable system also presents a large surface area of sample to the instrument. The illumination to the turntables is usually off-centre, so that sample cells from 15 to 25 cm down to scintillation vials can be used to scan samples of different sizes.

7.7.4 *Sample presentation for fibre-optic sensors*

Some instruments scan the sample directly, using a fibre-optic sensor to transmit spectral data from the sample to the instrument. The efficiency with which data can be passed depends on the wavelength range and on the distance from sample to instrument. Fibre-optics become less efficient the greater and longer the wavelength range. This is because there are more things that can interfere with the wavelengths at which the measurements should be made, and shorter wavelength ranges favour transmittance of spectra data. Between 700–1100 nm fibre-optic cables are effective over 10–12 metres. They are useful under operating conditions that are unfavourable to the instrument, such as at elevated temperatures. The size of the sampling head is quite small, and where this approach is used to scan static samples, several scans should be taken at different areas of the sample and the scans averaged to provide a more comprehensive image to the instrument. Wavelength ranges should be optimised for application of fibre-optics presentation systems to identify wavelengths where constituents such as moisture can introduce interferences. Sapphire windows are usually used with fibre-optics sensors.

On-line NIRS systems that continuously scan moving belts or trays of samples, mixers or other types of containers where processing is taking place, use no sample cells. The moving sample serves as the cell. These instruments overcome the potential errors of non-uniform surface by taking and averaging multiple scans. Modern diode-array NIR instruments take many scans per second over a wavelength range of 900–1700 nm. Filter instruments work over the small wavelength range produced by careful selection of discrete filters, and are equally applicable to on-line analysis. Samples of liquids can also be presented for NIRS analysis by on-line instruments by means of flow-through cells, where again many

scans are taken and averaged. Flow-through cells intended for use with slurries and liquids that may contain abrasive materials, such as sand, usually use sapphire windows which are very hard and resistant to abrasion.

Analysis of soil by NIRS has difficulties associated with sampling and sample preparation. Soil has to be sampled at more than a single depth to represent the composition of the soil through the profile of the growing depth. Preparation of the samples for reference analysis and NIRS scanning is further complicated by the fact that the top 10–15 cm or so are usually full of roots. The composition of the roots is quite different from that of the soil and will affect the spectral signals, and therefore also the results of the analysis of the soil component. Commonly-used soil profiles include 0–15 cm and 15–60 cm. Soil texture also varies even within a field, which further complicates sample preparation. The ultimate way of soil analysis is likely to become a probe, equipped with a NIR fibre-optic sensor that can be inserted into the type of hole made by a soil auger, and make direct measurements of the soil *in situ*. This will give a closer estimate of the soil moisture and texture as it is in the field. Subsequent analytical results determined on air-dried, ground soil should be converted to the as-is reference data for calibration purposes.

7.7.5 Sample presentation for interactance

Interactance (see Chapter 6, section 6.2.3) is a system where the instrument is brought into direct contact with the sample, using a special type of fibre-optic sensor. The sensing head carries a window, usually of sapphire, and the sensor is placed directly onto the surface of the material to be tested. It is particularly useful for testing changes that are transmitted through skin of animals including people, and can also be used for testing materials such as vegetables and large fruits, such as pineapples, which cannot be conveniently sampled, partly because of their high moisture content and partly because of their size, shape and surface characteristics. The interactance sensor can be applied to areas of the surface that have been shaved in several areas to expose a smooth surface for application of the sensor. Surface water can introduce a degree of non-linearity by increasing specular reflectance. Interactance scanning of high moisture samples, such as fruit and meat can reduce possible errors caused by surface water.

References

AACC. 2000. AACC International Approved Methods of Analysis. 11th Edition. Method 44-15.02. Moisture -- Air-Oven Methods. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-44-15.02.

8

Variables that can affect performance of NIR instruments

There is no such thing as an error-free operation. One rule that seems to work is that the sources of major errors are usually easy to locate and correct, because only a few factors can cause major errors. Failure of a lamp or an internal cooling fan can cause obvious changes in output from one sample to the next. Accidentally using the wrong calibration can also be a major source of error. Another source of such errors is that of mistakes in entering the reference data. Entering one item of reference data twice, or failure to enter it once will place all subsequent entries wrong by one position, which will disrupt the computing of the calibration model. Small errors, whether consistent or variable, are more difficult to find because they can be caused by so many factors.

NIRS, as an analytical tool, differs from chemical analysis in that it derives absolutely from the spectra. It is possible to obtain spectra of different spectral characteristics from samples with closely similar or even the same chemical composition as reported by reference testing. This is because of external influences on the material. This section is included because it is important to be aware of the sources of error in order to correct or compensate for them. Over 40 factors can affect the efficiency of NIRS testing. The main sources of error are

summarised in Table 8.1. Most of the factors under 'Sample' are associated with agricultural and food materials. Individual operations in other fields will identify their own specific factors.

Table 8.1 Main sources of error in NIRS applications

Source of error		
Instrument	**Sample**	**Operator**
Instrument design	Sample type (e.g. solid, liquid)	Reference method**
Instrument stability	Sample size, shape and colour	Sample preparation
Wavelength precision	Sample texture	Sample presentation
Non-linearity	Sample storage	Calibration development
Stray light	Chemical composition	Wavelength optimisation
Instrument-to-instrument variability	Growing location	Mathematical treatment
Instrument sensitivity*	Path length	Sample thickness
Wavelength range	Growing season	Operating conditions
Sample presentation system	Stage of growth	General carelessness
Sample cell variance	Sources of ingredients	
	Processing system	

*Sensitivity to temperature, relative humidity, dust and vibration
**Reference methods used in entering data, calibration development, evaluation and monitoring

8.1 Variables

Variables that can affect instrument performance include factors that come from the instrument itself, its design and especially its sample presentation system, the sample and the operator. Static electricity is a factor that is attributable to the instrument, the grinder and the sample. It causes the particles of ground material to become oriented in different patterns in the sample cell to the extent of affecting the signal, and the NIRS results. This phenomenon can be overcome by rotating the cell through 180° and taking two scans. Instruments that rotate the cell during scanning eliminate static electricity as a source of error.

8.2 Instrument factors

Basic instrument design includes the optical system, i.e. whether the instrument uses discrete filters, a monochromator, diode array, interferometer or some other form of wavelength range selection, the detector(s) and the sample presentation system. Stability in output is essential to consistent, accurate NIRS analysis. This is affected by several factors, including wear of moving components, mechanical vibration, aging and deterioration of components, dust build-up on components involved with optics, and expansion and contraction of components due to temperature fluctuation. Instruments with no moving parts in their electro-optical systems, such as diode array and CCD instruments, do not appear to be encumbered with this type of mechanical wear, but their sample presentation system may use a revolving turntable, the motor of which could generate noise due to wear.

The components of all types of instruments are affected by temperature fluctuations, and precautions must be taken during scanning material for calibration development to include the temperature range likely to be encountered during day-to-day operations. Typically, the output of optical components, such as discreet filters, can vary by about 0.5 nm for each change in temperature of 1.0°C. The influence of this on analysis depends on the material and factor to be analysed. For some applications a change of as little as 0.1 nm for a particular absorber can affect the results, while for others changes of even 2.0 nm can be tolerated. In development of the most robust instruments, manufacturers strive to anticipate interactions between components and temperature by employing internal temperature control, using fans and in-line thermo-electric controls. Failure of a fan can cause abrupt changes in performance.

Operators can compensate for these fluctuations during development of calibration models simply by scanning the samples throughout the working day. Before the working day begins the atmosphere of a laboratory has settled down from the previous day and is stable. As soon as the staff enter the laboratory they change the atmosphere by increasing the temperature, maybe by only 1 or 2 degrees, but that can have an effect. The staff also change the relative humidity by breathing out moisture vapour.

Dust is an ever-present hazard, and where possible, instruments should be freed of dust build-up by using equipment such as clean air sprays, vacuum and brushes. The action of scanning samples changes dust levels. However, by scanning samples for calibration development or day-to-day analysis during the normal

working day, the instrument will have automatically been exposed to all of these fluctuations, which will then have been built into the variance of the calibration.

Most NIR instruments are also sensitive to relative humidity within the instruments. If the humidity within the instrument changes during scanning, atmospheric absorption bands will show on the sample spectra. These very sharp bands, less than 2 nm in width, are most evident in the 1300–1400 nm and 1850–1900 nm regions. If the calibration makes use of data in these regions, the predicted results will be affected. Transmittance instruments use atmospheric air as a standard against which the sample is compared. Changes in the relative humidity can be expected to cause changes in the air used as a standard, from sample to sample. For this reason, NIT instruments should not be located too close to a doorway, open window or items such as a coffee maker.

Spectral precision (see Chapter 9, section 9.1.2) is the reproducibility of the spectral signals on scanning a sample, and is both an instrument and a sample factor that affects performance. It is a function of the optical system, the detectors, the sample presentation system and the material being analysed. Spectral precision should be determined for all materials, for which the instrument is expected to be used. For example, spectral precision for scanning wheat flour is significantly better than precision for scanning whole wheat grains. Similarly, spectral precision of dried ground forages will be quite different from that of the original fresh, chopped material. The spectral precision of any material will affect all future NIRS analysis of the material.

Stray light is one cause of non-linearity. Some stray light is light energy transmitted by the optical system at wavelengths that differ from the desired wavelengths. Another source of stray light is light energy that reaches the detectors without having passed through the sample. External light, such as room light, could also reach the detector directly. Instrument designers take precautions to provide adequate protection to minimise this source of error.

Non-linearity is an instrumental factor that is beyond complete control of the instrument manufacturer. It is mainly a function of the signal received by the detector(s), relative to the physical and chemical composition of the material. Specular reflectance is an important cause. It is not diffuse reflectance and carries no information. The effect of specular reflectance is small at wavelengths of low log 1/R, but increases as the log 1/R increases, which is not necessarily at higher wavelengths. In terms of measurement of log 1/R, consider log 1/R as consisting

of log $1/R_1$ + log $1/R_2$, where log $1/R_2$ is specular reflectance. In most cases log $1/R_1$ is well above 95% of the signal processed by the instrument software.

In general log 1/R tends to increase at higher wavelengths, so the effect of specular reflectance also tends to increase at higher wavelengths. So, at higher wavelengths the signal that reaches the detector (which the computer uses to convert the signal into composition) increases because of the specular reflectance. Because specular reflectance carries no information this causes an increasing degree of inaccuracy at the higher wavelengths, which is manifested in terms of non-linearity. Mathematical pretreatments of the log 1/R signal cannot correct completely for this type of non-linearity. This can also be affected by the characteristics of the material being scanned. Instrument non-linearity can be also caused by temperature changes within the instrument, by changes in the lamp and by mechanical changes within the instrument. Another source of non-linearity lies with materials that may exude liquids at the surface of sample cell windows when they are compressed during loading the sample cell. The exuded liquid changes the transmission capability of the window. This can be a serious source of non-linearity with the applications of NIRS to fresh materials, such as fruits and meat, where sample presentation is one of the most important aspects of the application.

The internal standards used in NIR reflectance instruments may differ and vary slightly in composition from instrument to instrument. This affects the efficiency of calibration transfer among instruments of the same make and model. Air is the standard used in transmittance instruments. Theoretically, this should contribute to more efficient transferability between NIT instruments than NIR reflectance instruments. In most NIR reflectance and transmittance instruments, the sample is viewed through a built-in transparent window, which is not part of the sample cell. The windows may be made of glass or quartz. The composition of both types of window is subject to variability. This has been discussed and illustrated in Chapter 6 (see section 6.3.8).

For practical purposes, by the time an instrument is offered for sale, any technical problems have been carefully evaluated and solved by the manufacturer. Modern NIR reflectance and NIT instruments carry diagnostic software that enables the operator to monitor their performance. The diagnostics, or at least a check cell, should ideally be run daily at staggered intervals, to check whether changes take place during operation. The presence of people and other environmental factors cause fluctuations in temperature, relative humidity and dust. Records of the results of check cells and diagnostics should be maintained, for long-term monitoring

of the instrument. Check cells are useful in monitoring day-to-day instrument performance. These are prepared from the commodity or commodities with which operators will work. The cells are sealed to prevent changes in the surface. Scanning the check cell 2 or 3 times during the day acts as a back-up and is faster than running the diagnostics. Provided the check cell results are consistent from day to day, the diagnostics can be run once or twice a week. Maintaining a record of all check cell and diagnostic data is helpful to the instrument manufacturer whenever repairs are required.

Instruments used outdoors are exposed to wide fluctuations in conditions. Temperature, the relative humidity of the working environment, dust and other extraneous factors all exert their influences on instrument components and performance. Effective performance has been obtained from instruments operating under field conditions with temperature ranges from 2°C to over 40°C. Calibrations should ideally be developed over a period of time that exposes the instrument to the most likely fluctuations in these factors. Instrument performance should be monitored as closely as practicable under all operational conditions.

One of the reasons underlying the proven success of the ANN calibrations offered by some instrument companies is that the spectral and associated reference data have been assembled from a very large number of instruments. The data collection also represents an impressive range of fluctuations in working environment, from the point of view of temperature, RH, dust and reference laboratory performance. This assembly of variance acts as a stabilising influence on calibrations. Yet another factor is the altitude and barometric pressure, which in turn influence the relative humidity of a work area.

The diversity of materials analysed by NIRS calls for ingenuity in design of the sample presentation system. The fundamental goal of a sample presentation system is to present the sample to the instrument the way it really is and in a consistent manner. Some instruments use an open sample cup that rotates on a turntable. To get the most consistent results the surface can be leveled by ‘striking’. A better system is the use of a cup based with a transparent window, with the sample viewed (scanned) from below. This presents a uniform surface and eliminates potential error caused by not leveling the sample, or by differences in leveling technique. For such cells the window is made of special glass, quartz or sapphire.

The sample presentation system can become an important source of error. The extent of this error can be determined by reloading and scanning a check sample

several times and predicting the results. The SD (standard deviation) of these is compared to the overall SEP (standard error of prediction). The percentage of error attributable to the error of the NIRS test can be calculated as follows:

1. Square the SEP ...A
2. Square the SD of the check sample results ...B
3. Subtract $(SD_{checks})^2$ from the $(SEP)^2$ ...A − B
4. Extract the square root of (A − B) and divide by the SEP ...C
5. 100 − (100 − (C × 100)) = % error attributable to the NIRS test

Example:

If,

$$SEP = 0.250;\ SD_{checks} = 0.198$$

Then,

$$A = 0.0625;\ B = 0.0392$$

$$C = \sqrt{(0.0625 - 0.0392)} / 0.250 = 0.153 / 0.250 = 0.611$$

Error attributable to the NIRS test = 100 − (0.611 × 100) = 38.9%

The NIRS test error includes instrument error, as well as that of sample presentation. Table 8.2 gives some hypothetical values for the NIRS test error as a percentage of the overall SEP. The data of the above example are for testing wheat for whole-grain protein content, but are applicable to any NIRS test. The SD of check samples includes the error caused by sample preparation and presentation. The different values are included to illustrate their influence as a proportion of the total SEP. The protein data show, for example, that if the error of check sample analysis is about halved (from 0.198 to 0.107, as measured by SD of the results of prediction), the error caused by sample presentation (including instrument error) is reduced by about 75%.

Some new and prototype instruments have been found to include sample presentation systems that were responsible for over 70% of the overall SEP. The importance of using a check sample to provide the reproducibility information where practicable, cannot be over-emphasised. When the error induced by sample preparation and presentation is low, relative to the overall SEP, it is important to research the other sources of error, in order to improve the SEP.

Table 8.2 Influence of error attributable to the NIRS test on overall SEP

Constituent	Protein content (%) of wheat	
SEP including SPP*	0.250	Overall SEP
SD of Checks	0.198	39.0
	0.149	19.7
	0.107	9.6
	0.085	6.0

*Sample preparation and presentation

Path length, in terms of NIR spectroscopy, is the distance that radiation travels within the sample, and is important in both reflectance and transmittance technology (see Chapter 6, section 6.2.1). As a variable it becomes more a factor of the sample than of the instrument. In transmittance, the sample thickness (sometimes referred to as 'path length', but not to be confused with the actual path length) should be optimised during calibration development. Sample thickness should be optimised by any operator who wants to develop a calibration for a new commodity. This is done by developing and evaluating calibrations with the same sample sets of at least 50 samples, but with different sample thicknesses, e.g. from 5 to 30 mm. The thickness that gives the best statistics is the one to use.

Sample presentation system design includes sample cells. In reflectance mode, the external optical path length is essentially constant but it has been mentioned in previous sections that differences occur among sample cell windows in both composition and thickness. These can both introduce biases in analysis (Williams, 2006). Sample cell windows can vary in thickness and are not always completely planar. If the window is of different thickness, the diffuse reflectance surface presented to the detector will vary by up to several micrometers. In reflectance instruments that use sample cells, these small differences change the path length in parts of the surface, and can cause biases. Biases introduced in this manner can be determined by rotating the sample cell through 180°. Instruments that rotate the sample cell, eliminate this source of error.

Sample cells may differ in spectral precision from one another. Operators are advised to buy more than one cell, to compare the spectral data for different commodities for each cell, to compare the performance from cell to cell and to determine whether differences occur among cells. If such differences are noted, two or more cells should be used in calibration development. This minimises the occurrence of biases whenever the different cells are used. In the absence of this, if the cells are numbered, and any of the cells is associated with a small, but

consistent bias, the bias could be corrected for when such a cell has to be used, e.g. due to breakage of other cells. Differences among sample cells are intensified among cells used in automated sample presentation where trays of samples are presented to be scanned sequentially. Variation of over 0.02 absorbance units have been detected among sets of sample cells which were ostensibly of the same characteristics. This type of variance should be detected before, and included in calibration development because it can result in significant differences in predicted results.

8.3 Sample factors

These include the basic constitution of the sample, such as solid, liquid or slurry. The basic constitution includes whether the sample contains foreign material, such as weed seeds, chaff, or broken seeds in grain, stones in soil, solid inclusions in manure slurries, shape, size, texture (degree of hardness or softness) and colour. It also includes particle size and bulk density of ground materials and powders, viscosity of slurries, the presence of weed species in forages and other factors. All of these items affect spectral characteristics, and contribute to diffuse reflectance and to the variance that has to be included in the calibration. In many industrial applications, the material to be tested is fairly consistent in its basic form, such as grain or forage. In the application of NIRS, feed manufacturers face a more complex situation because of the diversity of the ingredients used in feed mixes. Growing location and season both affect all crops, including grains, forages and others. Important sources of variance in forages include the species of grasses and other types of plants that make up the populations and the stage of growth. The physical and chemical composition of grasses are different from those of legumes and other types of plants. For example, the stems of lucerne (alfalfa) are more 'woody' than those of grasses, and become fibrous more quickly than grasses during maturation.

Applications to fresh forages are complicated by the sizes of the plants, and the need for a consistent way to present them to the instrument. Particle size, shape and bulk density affect packing in sample cells or scanning compartments, which affects the diffuse reflectance. Table 8.3 gives the bulk densities of some common powdered commodities, ground in a cyclone grinder fitted with a 1.0 mm screen. The figures are the means of triplicate tests carried out using a stoppered measuring cylinder. Sample weight was 25 g for all materials except for dried alfalfa (20 g) and barley straw (10 g). The cylinder was filled using a stem-less funnel, and gently tapped twice before the measurements of volume were taken.

Table 8.3 Bulk density of some powdered materials

Material	Bulk density (g/cm^3)
Faba bean	0.64
Lentil	0.54
Wheat flour (hard wheat)	0.51
Maize (corn)	0.51
Wheat (hard)	0.50
Soybean hulls	0.49
Barley	0.39
Wheat bran	0.35
Oats (whole)	0.34
Defatted canola seed	0.34
Dried alfalfa (lucerne)	0.24
Barley straw	0.14

Kernel texture affects the particle size of ground grains. In the case of whole grains, texture is a factor of the association between adjacent cells, mainly of the endosperm, which forms the bulk of cereal grains. Differences in the spatial arrangement of cells in the endosperm affect the diffuse reflectance of light energy, which makes it possible to predict changes in kernel texture in intact kernels by NIRS. Some radiated energy passes through the kernels (or seeds), and the spectral signals are a combination of the energy passing through and being diffusely reflected from the surfaces of the kernels. Figure 8.1 shows the average spectra of three composite samples of ground and whole wheat of different kernel texture.

Kernel texture is affected by factors, such as genetic make-up (commercial varieties of grains differ in texture), growing location and season, chemical composition, moisture content, soil fertility and cultivation practice, harvest and storage conditions and others. Growing location and season both have strong effects, and calibrations for the large-scale industrial NIRS analysis of agricultural crops should include samples from several growing seasons.

In the case of NIRS analysis of whole seeds, seed or kernel size influences the diffuse reflectance of whole seeds and has an impact similar to particle size in the ground seeds. Figure 8.2 shows the effect of kernel size as indicated by

1000-kernel weight (TKW) on the spectra of hard red spring wheat. The spectrum of the wheat with biggest kernel size is uppermost. The samples of high, medium and low kernel weight were obtained by sieving the sample through sieves with oval apertures of 2.8, 2.5 and 2.2 mm.

Sources of wheat intake for flour mills and of feed ingredients for feed mills constitute the most important sources of variability in the materials with which they work.

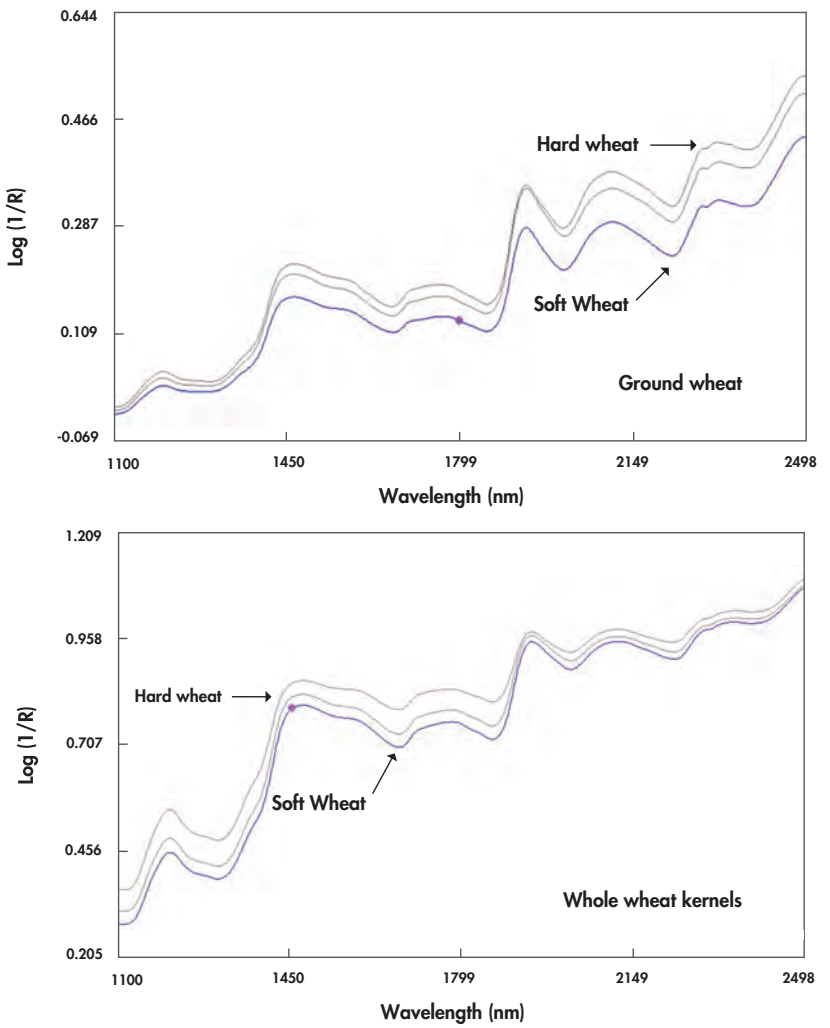


Figure 8.1 Influence of kernel texture on NIR spectra of ground wheat and whole wheat kernels.

Chemical composition affects the texture and spectral characteristics of most materials. Protein, oil, fibre and particularly moisture contents all exert a strong influence on the spectral composition of materials. They influence kernel and

seed texture, as well as the texture and characteristics of forages and straws. High oil content causes grains and seeds to paste during grinding, which affects the diffuse reflectance. Factors that change the particle characteristics of ground samples cause changes in the path length (see Chapter 6, section 6.2.2). All of these sample factors are the reasons why application of NIRS to agricultural and food materials is more complicated than applications to many industrial materials, where texture is not so strongly affected by interactions among the constituents. Some of these factors can be overcome by mathematical pretreatments.

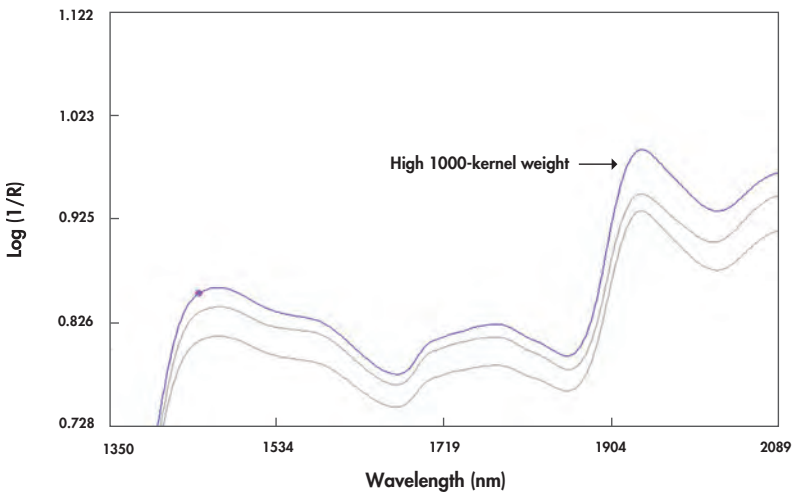


Figure 8.2 Influence of kernel size on spectra of CWRS wheat.

Moisture content can be considered as an error source. Figure 8.3 shows the influence of moisture content on the spectra of whole maize (corn). An NIR instrument senses the sample as it is. The results of reference testing are used in development of calibration models, and unless these results are reported on the basis of the original moisture content of the samples (as scanned by NIRS) the reference data will not truly represent the spectral data. This can be overcome up to a point by calibrating directly to reference data that have been corrected to a constant moisture level, but this will not be effective if the moisture range is very high, such as greater than 5% (e.g. from 10.0 to 15.0%).

Figure 8.4 shows the influence of water content on the spectrum of fresh chicken breast. Moisture content of the chicken breast was about 85%. The spectra showed no significant information above 1900 nm in the log 1/R or 2nd derivative (2 4 4 1) spectra (gap and segment of 4 wavelength points) because the detector became 'saturated'. The 2nd derivative showed that above 1900 nm the 'spectrum' became essentially noise.

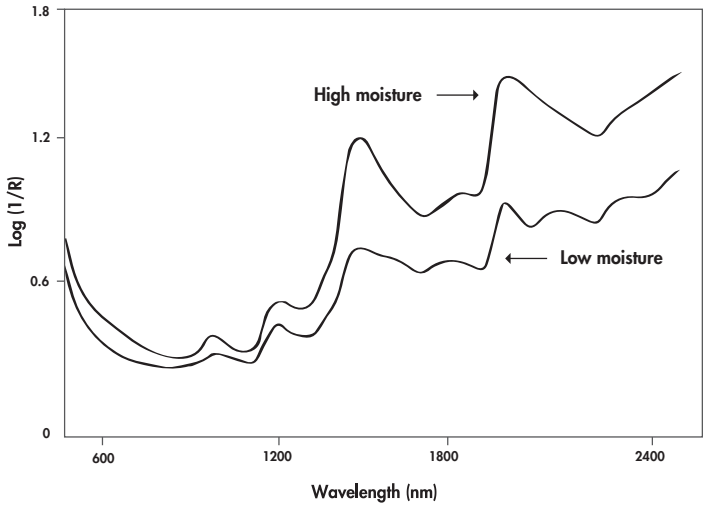


Figure 8.3 Spectra of high (54%) and low (12%) moisture maize.

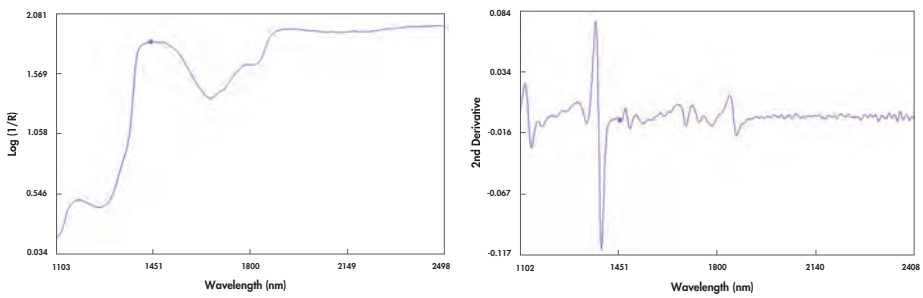


Figure 8.4 Log 1/R spectrum of fresh chicken breast and the 2nd derivative (2 4 4 1) spectrum of the same sample (gap and segment of 4 wavelength points).

Figure 8.5 illustrates the influence of growing location and season on wheat spectra. Each spectrum in Figure 8.5 shows the average of about 120 samples of CWRS (Canada Western Red Spring) wheat grown on 10 locations in western Canada in 2000. Growing season seems to exert more influence than growing location within a season (Figure 8.6).

Another type of non-linearity can be met with prediction of some functional factors. Functional parameters may not be closely related to composition. For example, no specific absorbers are associated with wheat kernel texture. Functional parameters are more likely to be associated with physicochemical, rather than simply chemical factors. So, this type of non-linearity may be more pronounced in the prediction of functional parameters, particularly those that are related to differences in texture.

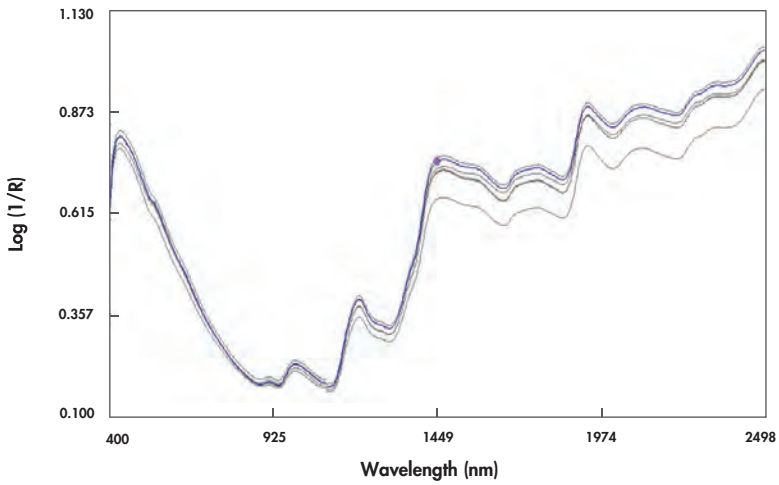


Figure 8.5 Influence of growing location on spectra of CWRS wheat grown at ten locations during one season.

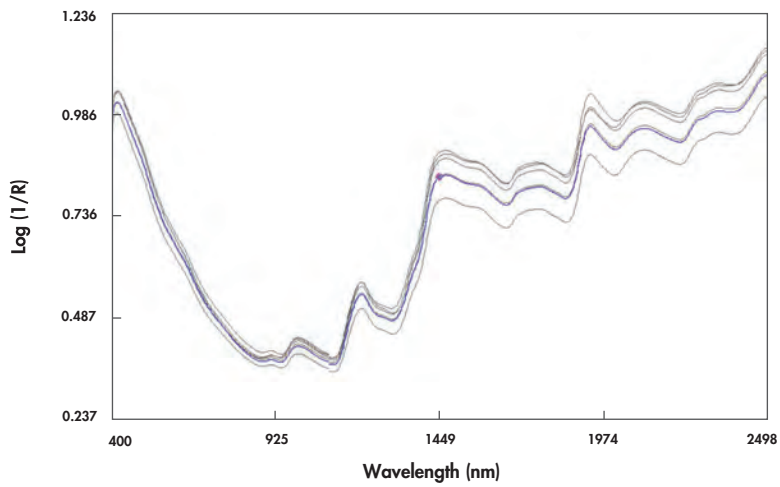


Figure 8.6 Influence of growing season on spectra of CWRS wheat grown over seven seasons.

Species and stage of growth are the main factors affecting spectral characteristics of forages. For the most part, straws are from mature grains, but forages are grazed and harvested for analysis from early stages. The main changes in composition are in protein, moisture, and fibre contents. Lignification develops as the plants mature. This has the effect of stiffening straws, and changes the particle characteristics of ground material.

Viscosity, uniformity in composition and total solids content are the most important factors that affect slurries such as liquid manures. Figure 8.7 shows log

1/R spectra of thin and viscous samples of liquid hog manure. Note the leveling-off of the thin spectrum above the 1900 nm water band area. The 2nd derivative of these spectra showed no absorbance bands in this region.

The thin manure contained over 99% water and the viscous manure contained 88% water. Slurries, such as liquid manure, introduce another type of complexity. When scanned in transflectance mode, the particles tend to agglomerate as the total solids increase. The radiation (light) may not penetrate the sample completely, and then transflectance becomes diffuse reflectance (see Chapter 7, Figure 7.7.2). The same phenomenon can happen with milk of low and high fat content. Fat content of milk can vary from less than 1% to over 10%. Both types of liquid are subject to sedimentation. Particles of liquid manure sediment downwards, while the fat globules of milk sediment upwards. It is important that such liquids are agitated during scanning in order to minimise the effects of sedimentation. Continuous on-line analysis ensures such agitation.

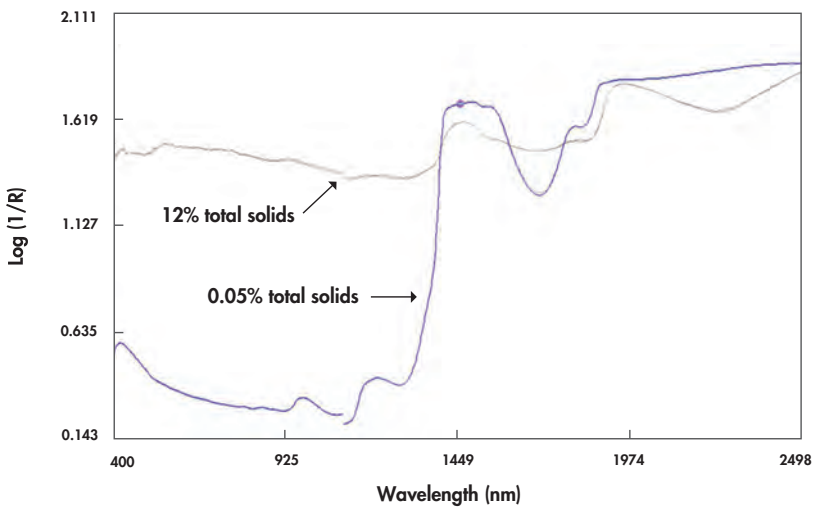


Figure 8.7 Log 1/R spectra of thin (0.05% total solids) and viscous (12% total solids) liquid hog manure.

Soils present a further set of variables. Of these, mechanical composition and moisture content rank foremost. Soils with high proportions of clay tend to be intractable and form dense agglomerates. They also tend to adsorb water to a greater extent than sandy soils. Removal of moisture and size reduction are essential to accurate analysis, but like other materials, soils should ideally be scanned and analysed by NIRS in their natural state before, as well as after drying and size reduction.

Because the composition and texture of soils differ depending on their profiles, soils should be sampled and analysed at different depths. The uppermost 15 cm are usually full of roots and partially decayed plant material. This modifies both the composition and texture, and affects sample preparation. The efficiency of soil analysis has suffered in the past because of the labour and expense of carrying out sufficient sampling. For example, in an undulating field of 50–100 hectares, because of its flexibility and low cost per test, NIRS can revolutionise the costs of soil testing. Instrument attachments can be designed to enable direct, on-the-spot testing for moisture, nitrogen, carbon and other constituents at far more numerous sampling sites than reference analysis would permit. Due to of the speed and flexibility of NIRS, the efficiency of the testing can also be improved because all samples can be tested on-site, and there will be no need for compositing the samples to save the cost of testing. Areas of different fertility can be identified within a field, which can improve the efficiency of fertilizer use.

Most NIR reflectance or NIT users who have to deal with high moisture materials, such as fruits, vegetables, meat and fish work with the instrument companies to develop sample presentation systems. The high moisture and fat contents of meat and fish affect sample cell loading and clean-up. Between-sample clean-up is also a factor in scanning liquid manure, milk and other types of thin and greasy or gelatinous slurries. The method of sample storage affects the physical and chemical make-up of samples, and as a result it affects the diffuse reflectance and effectiveness of NIRS analysis.

Since its adoption as an authenticated analytical system, NIRS has found applications in an assortment of industrial fields other than those associated with agriculture and foods. The physical and operational factors entailed in manufacture exert influences on the spectra that are specific to the product. Staff engaged in different industries will recognise the variables specific to their operation, and take appropriate measures to accommodate them.

8.4 Operator factors

Before starting to work on NIRS for the first time, it is important to establish what reference methods are to be used and to determine the standard error of each method. The reference methods are what will be used for model (calibration) development and evaluation, and subsequent monitoring of the performance. The overall accuracy and reproducibility of an instrument includes the error of both NIRS and reference testing. The error of the reference method is the basis

from which one has to work, and its determination is chiefly the responsibility of the operator.

Sampling, especially sub-sampling, and sample preparation are key factors in successful analysis by any method. One of the difficulties with NIRS analysis of whole commodities, such as grains and forages is that while NIRS can be applied to the sample as it is received, reference testing almost invariably calls for sample preparation, and usually sub-sampling. For example, a sample of over a kilogram can be analysed for protein content by an NIR instrument, but the reference test has to be performed on a gram or less of dried and ground sample (equivalent to only about 28 kernels of wheat or barley).

Grinding the sample during sample preparation results in losses in moisture content that vary depending on the original moisture content, the type of grinder and the number of samples ground during a single session. For accurate analysis by combustion, the sample should be ground to pass through a 1.0 mm screen in the grinder. Moisture loss for some grains during grinding by the Cyclone grinder is summarised in Table 8.4.

Table 8.4 Moisture loss during grinding bread wheat, durum wheat and barley with a Cyclone grinder

Original moisture (%)	Moisture loss (%)		
	Wheat	**Durum**	**Barley**
17	3.2	3.0	4.0
15	3.2	2.6	3.5
13	1.5	1.4	2.5
11	0.9	1.1	1.9
9	0.3	0.5	0.8

Losses using a burr-type mill are about one half of the values shown in Table 8.4, but the error of test methods such as combustion protein analysis is increased due to the changes in particle characteristics of the ground sample and the smaller sample size used in the testing. Burr mill-grinding is more sensitive to moisture content than cyclone-grinding in terms of particle characteristics. Another factor for which operators are responsible is grinder maintenance. Wear of items, such as the screens in grinders and corundum discs causes changes in particle size, bulk density and packing in NIR instrument cells. Changes in the rpm (revolutions per

minute) of grinders caused by ageing of the motors also cause changes in particle size. This also becomes important in reference tests. It is a good idea to assign a grinder specifically to grinding for NIRS testing, and to monitor its performance. For laboratories that work with paddy rice, the hulls are extremely abrasive and cause wear of grinding surfaces and screens of cyclone grinders. This changes the particle size of the ground material.

An error of 1.0% moisture causes errors of about 0.15% protein in wheat, approximately 0.2% in canola seed and about 0.4% in soybeans. The cumulative error in testing the whole and ground grains for moisture content can easily exceed 1.0%. Errors of even 0.1% can become expensive when a difference of 0.1% in protein content can mean a dollar per tonne. To protect against arguments over prices based on analysis, the combined errors must be evaluated and built into contracts. For example, in western Canada tolerances of 0.4% are accepted by both farmer and elevator manager before compensation has to be paid. This is related to the 95% confidence limits of the precision of the protein test (approximately ±0.2%). In some countries the tolerances are even higher.

Undetected moisture loss is an aspect of sample preparation. Oilseeds and other seeds, the results of which are reported on a moisture-free (dry) basis, are usually dried before analysis by reference methods since the moisture interferes with the determination of oil by extraction. If the seeds are dried whole and ground for analysis, the ground material will absorb water rapidly, and may contain 2 to 4% moisture by the time it is analysed. This undetected moisture can induce significant errors to the reporting of reference results for oil or protein. Figure 8.8 illustrates the difference in the moisture band at 1930 nm for soybeans dried before and after grinding. Note the commensurate increase in the protein band at 1978 nm, as the moisture band at 1930 nm decreases as a result of the different drying processes.

The moisture content of the samples of ‘dried’ soybean seed varied up to 3 to 4%, enough to cause errors of up to 2% in protein content. The reason for the absorption of water is the presence of very hygroscopic cellulosic fiber in the oilseed testa-pericarp. If the oilseeds are to be analysed by NIRS there is no need for drying the samples before analysis. Moisture can be determined directly by NIRS to enable reporting of the oil and protein results on a moisture-free basis.

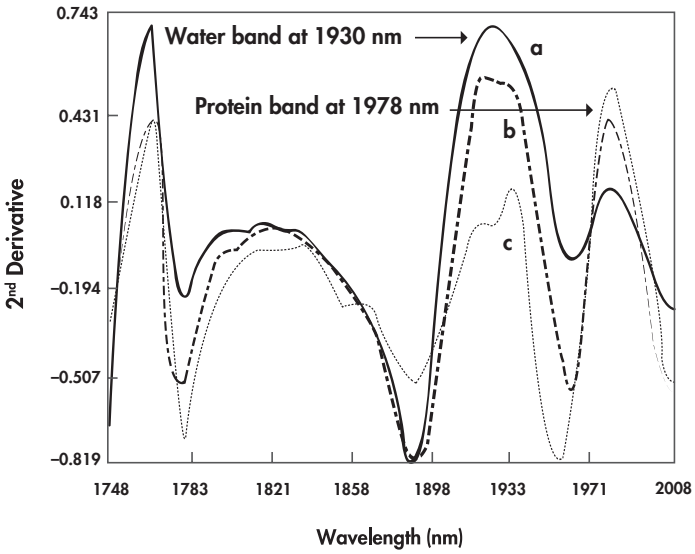


Figure 8.8 NIRS spectra of soybeans: a) whole as-is; b) after drying and grinding; and c) after grinding and drying.

Sample presentation on the part of the operator calls for care and attention, and general efficiency in presentation of the sample to the instrument. For most whole grain analysers, this has been simplified by use of the hopper-type sample access. For operations that use a sample cell, whether it is closed or open, some skill is involved with filling the sample cell. With instruments that use an open glass-bottomed cell where the sample is viewed from below, cell loading is simplified, but the operator has to make sure that the cell contains sufficient thickness of sample – for most grains, at least 3 cm is recommended. Large grains, such as maize and sunflower seed require a thickness of 5 cm.

Operating conditions include temperature, relative humidity and dust levels. As a general rule, if the conditions are comfortable for the operator they will be suitable for the instrument. In areas like grain elevators, flour and feed mills, and other industrial situations it is difficult to provide more than reasonable control over these factors.

Calibrations are always a possible source of error in NIRS testing. Calibration methods are the responsibility of the operator, and unless the development of the calibration model is carried out properly it will be impossible to obtain reliable results by NIRS testing. If the calibrations have been purchased from an instrument company or a software company, they have to be tested by the operators. This is done with samples of the materials for which they intend to

use the instrument. Accurate reference data should be used to verify that the calibrations are applicable with no biases. Calibration development is discussed in detail in Chapter 9.

Carelessness is a hazard in any repetitious operations. It can cause unreliable results and, in some cases, it is responsible for accidents. The main reasons for carelessness are boredom (in routine analysis) and inadequate allocation of time. The time factor is caused by operators trying to work too fast. This causes mistakes, such as failure to record results properly and spillage of samples. Telephone calls and other interruptions at inconvenient times are frequent causes of excessive speed. Stimulating operator interest in the purpose of the testing can alleviate the boredom problem to a certain extent. The working day consists of twenty-four hours, of which eight can be assigned to sleeping. The remaining time is divided between the actual work, including transport to and from work, and the living part of the day. It is useful for supervisors of laboratories and other operations to bear in mind that their staff are all people with their own personal issues, which may intrude upon their concentration on the job. Attributes of successful supervisors include degrees of patience and tolerance, in addition to their personal skills and thorough understanding of the operation.

Reference

Williams, P. 2006. The near infrared window – glass or quartz? A study in precision. *Journal of Near Infrared Spectroscopy*, 14, 127-138.

6

Calibration development and evaluation methods

Calibration has been the main challenge to applications of NIRS since the introduction of the technology almost 45 years ago. Several companies now supply instruments with factory-developed calibrations for an array of commodities and constituents. The calibrations can also be purchased subsequent to purchase of the original instrument. The companies can develop calibrations using samples supplied by the potential client or by inviting clients already using their NIR instruments in order to supply them with spectra and reference data – the system originated by Tecator for the development of their ANN calibrations for wheat protein and moisture contents some 30 years ago. Nevertheless, calibration is still needed for any new applications (e.g. soils, manures, fresh materials, such as forages, meat, fish and a host of industrial applications), as well as for updating existing calibrations for new sources of variation, whether the models are developed by the client or the factory. This chapter provides guidelines for calibration development and evaluation methods applicable to any commodity, constituent or parameter.

An important message to all NIRS users is *look at the data* (spectral and reference). Ideally, this should be done as the data are being generated during daily analysis,

or while recording spectra for development of the calibration model. In this way the operator will become familiar with what the data should look like, and also the shape of the spectra. The shape of the spectra will differ depending on the wavelength range for the same sample, as illustrated in Figure 9.1. All spectra are from the same sample. The shapes of 1st and 2nd derivative spectra (see Chapter 5, section 5.2.5) are very characteristic.

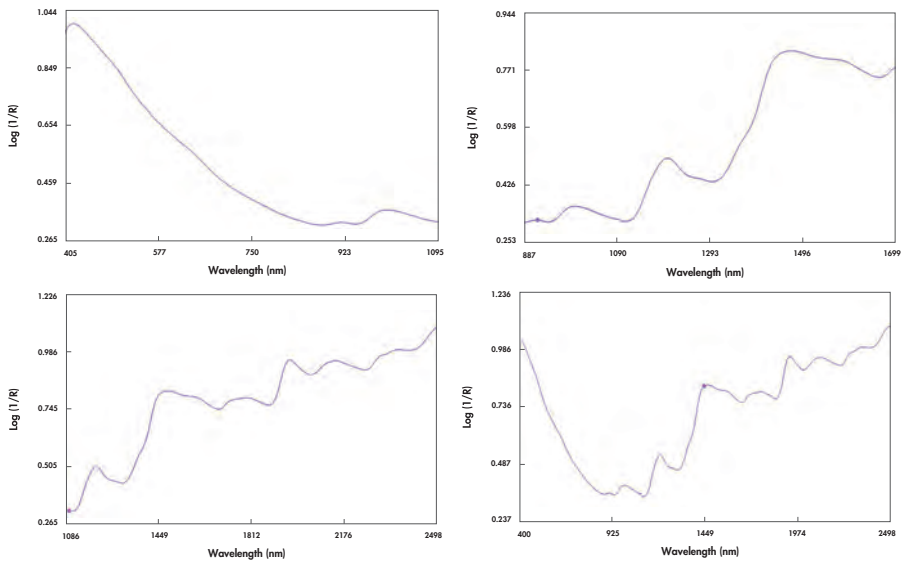


Figure 9.1 Log 1/R spectra of whole wheat at different wavelength ranges: 400–1100 nm; 900–1700 nm; 1100–2500 nm; and 400–2500 nm.

9.1 Basic steps in calibration

Access to comprehensive sample sets with reliable reference data is absolutely essential to calibration model development. Modern chemometrics have greatly streamlined and simplified calibration model development. However, no matter how sophisticated the chemometrics, they cannot compensate for poor sample set assembly, i.e. failure to include enough samples with proficient sampling technique that includes all of the anticipated spectral variance, and physical and chemical composition.

Whether a calibration is developed by the client or the factory, there are a number of basic steps to establish a reliable model. These are essentially the same for any application, and are summarised in Table 9.1. The steps are in sequential order and not in order of importance – they are all important.

Table 9.1 Steps in the development of a calibration model

Step 1	Determine or develop the methods to be used as reference for calibration and monitoring
Step 2	Determine the standard error of each reference method (SET or SEL)
Step 3	Determine the precision of the spectral data with the materials you will be analysing
Step 4	Identify all sources of variance for the materials you will be analysing
Step 5	Assemble samples that accommodate all sources of variance, ideally replicated
Step 6	Identify or develop a sample preparation system (for reference and NIRS analysis)
Step 7	Prepare samples for reference and NIRS (if necessary) analysis
Step 8	Identify or develop a sample presentation system for NIRS analysis
Step 9	Develop a repeatability file – optional (depending on software), but very useful
Step 10	Scan samples and view the spectra
Step 11	Select samples for calibration/validation on basis of reference or spectral data
Step 12	Perform reference analysis and add to spectral data
Step 13	Develop calibration model
Step 14	Evaluate and optimise calibration model, including wavelength range and mathematical pretreatment (or preprocessing)
Step 15	Enter calibration model into the NIR instrument
Step 16	Verify the precision (reproducibility) of the NIRS analysis
Step 17	Verify accuracy of the NIRS calibration model by analysis of completely new samples
Step 18	Carry out slope/bias corrections where necessary
Step 19	Re-analyse fresh samples to verify accuracy has been restored
Step 20	If a network is to be developed, test calibration transferability among instruments

9.1.1 Steps 1 and 2 – reference methods and their error

The first step is to identify the reference method, and secondly to determine its standard error (see Chapter 4, section 4.1.9). The SD (standard deviation) of the reference test (SET) is sometimes referred to as the SEL (standard error of the laboratory). For some applications it may be necessary to develop a reference

method and verify its effectiveness in terms of providing the desired information about the sample before progressing to Step 2. The reference method is the basis of the use of NIRS to testing any commodity, constituent or functional parameter. It is used to develop and evaluate the calibration model and in subsequent appraisal of the calibration performance, by checking the accuracy of subsequent analysis. Every aspect of the application relates to the reference method, and spectral data, and the establishment of their integrity is fundamental.

9.1.2 Step 3 – determine the precision of the spectral data with the materials to be analysed

No matter what chemometrics are employed in the development of a calibration, the instrument will apply the calibration model as a constant to any new spectrum that it records, and will translate the spectral data into terms of the reference data for which the calibration has been developed. Hence the efficiency of the NIRS predicted data depends heavily on the quality and reproducibility of the spectra. This step is especially important in the case of modern handheld instruments for use in the field.

No calibration is needed for this step because it is only concerned with the spectral data. It is carried out by scanning a thoroughly-blended sample 10 times after sample preparation, with reblending of the sample and reloading the sample cell between scans. The SD of the spectral data is then determined at selected wavelengths. Suggested wavelengths are 1210 nm (a -CH_2 band – 8264 cm^{-1}) and 2230 nm (a 'reference' band – 4484 cm^{-1}).

Spectral precision is determined using raw spectral data, and should be done before development of a calibration model for a new application. No calibration is required for its determination. It can be determined by scanning one or more typical samples of the material 10 times, and recording the spectral data at 2 or more wavelengths. The sample should be reloaded into the sample cell between each scan, and the recently-scanned sub-sample should not be returned to the original sample. After the 10th scan the sample should be left in the cell and re-scanned 9 more times without disturbing the cell. The SD of the spectral log 1/R data from the first 10 scans is the variability in spectral data that will be experienced during NIRS analysis of the material (reproducibility). The SD of the spectral data from the 10th–19th scans is the repeatability of the scanning, and is strictly a function of the instrument because there will have been no variance contributed by the operator. Table 9.2 gives typical data for spectral precision of whole wheat kernels at two wavelength points.

Table 9.2 Spectral precision (Log 1/R or Absorbance values) of whole wheat kernels

Scan	1210 nm*	Scan	1210 nm**	Scan	2230 nm*	Scan	2230 nm**
1	0.4741	**11**	0.4666	**1**	0.9046	**11**	0.8738
2	0.4641	**12**	0.4633	**2**	0.8871	**12**	0.8852
3	0.4708	**13**	0.4702	**3**	0.9111	**13**	0.8758
4	0.4549	**14**	0.4641	**4**	0.8923	**14**	0.8711
5	0.4802	**15**	0.4687	**5**	0.8947	**15**	0.8773
6	0.4733	**16**	0.4674	**6**	0.8684	**16**	0.8792
7	0.4691	**17**	0.4638	**7**	0.8791	**17**	0.8823
8	0.4744	**18**	0.4683	**8**	0.8663	**18**	0.8738
9	0.4723	**19**	0.4662	**9**	0.8905	**19**	0.8751
10	0.4666	**20**	0.4675	**10**	0.8738	**20**	0.8756
Mean	0.46998	**Mean**	0.46661	**Mean**	0.88679	**Mean**	0.87691
SD	0.00693	**SD**	0.00228	**SD**	0.01491	**SD**	0.00425
CV (%)	1.47	**CV (%)**	0.49	**CV (%)**	1.68	**CV (%)**	0.48

*samples scanned 10 times with reblending and reloading of the sample between scans
**the 10th sample scan 10 times without being disturbed

Spectral precision should be determined for application of NIRS to any type of material. Table 9.3 shows results of precision checks on two instruments using whole wheat, barley grains and wheat flour. The SD of the spectral data tends to increase at longer wavelengths. Note the excellent precision of the spectral data for flour, and also that the spectral data from the FT-NIR instrument were more precise than were those of the dispersive instrument using the same whole wheat sample. The wavelengths shown Table 9.3 are 'traditionally' associated with some constituents. The CV (coefficient of variation) can be expected to increase for more complex materials such as soil and forages, and especially for fresh and high moisture materials. But its determination is a useful contribution to understanding the statistics resulting from application of NIRS.

Table 9.3 Precision of spectral data obtained from a dispersive and a Fourier Transform NIR instrument

Constituent 'assignment'		Wavelength (nm)*					
		964	1154	1200	1696	2346	2346
		Water	Water	Starch	Protein	Reference	Oil
Statistic	**Material**	**Dispersive instrument**					
Absorbance	Wheat	0.3071	0.3730	0.4744	0.7507	0.9030	0.9881
SD		0.00351	0.00477	0.00631	0.0119	0.01291	0.01245
CV (%)		1.14	1.28	1.33	1.59	1.43	1.26
Absorbance	Barley	0.1284	0.1703	0.2326	0.4100	0.5684	0.6725
SD		0.00402	0.00334	0.00365	0.00517	0.00625	0.00686
CV (%)		3.13	1.96	1.57	1.26	1.10	1.02
Absorbance	Flour	0.0715	0.0689	0.1012	0.2596	0.4406	0.5566
SD		0.00055	0.00045	0.00052	0.00090	0.00145	0.00178
CV (%)		0.77	0.65	0.51	0.35	0.33	0.32
Statistic	**Material**	**FT-NIR instrument**					
Absorbance	Wheat	0.5750	0.6537	0.7593	1.0095	1.1756	1.2410
SD		0.00310	0.0034	0.00494	0.00777	0.00799	0.00782
CV (%)		0.54	0.52	0.65	0.77	0.68	0.63

*The FT-NIR instrument records data in wavenumbers (cm^{-1}). Wavelengths are nearest equivalents to corresponding cm^{-1}-values; Absorbance = spectral data

9.1.3 Step 4 – sources of variance

Establishment of a database is fundamental to practical application of NIRS in any area of industry. The database is originally established for calibration development, but is a very valuable resource in all future applications. Important is to determine the sources of variance. Many of these have been identified and discussed in Chapter 8, but others that are unique to a specific application may occur. Sources of variance affecting the spectra include changes in the testing environment. Local conditions include altitude. Industrial conditions include daily fluctuations in temperature, RH (relative humidity), dust intensity and barometric pressure. Scanning of samples for development of calibration models should be carried out under all of the conditions expected to arise during normal operations, whether in the laboratory or under industrial or out-of-doors conditions.

The sources of variance most often considered concern composition, but any factors that affect spectral characteristics, such as sample texture, bulk density and operating conditions are always key factors. Texture and bulk density affect surfaces of materials such as paper. This influences the diffuse reflectance and therefore the interaction between sample and instrument. Any factors that represent changes in processing conditions that can change the diffuse reflectance of the sample should be included. This applies to solid, liquid, or viscous samples. If samples of grain are to be ground before scanning, the temperature will increase during grinding. Under daily operating conditions it is usually not practicable to allow them to cool before testing, so ideally the samples used in development of the calibration should also be ground and scanned right away. If the samples are ground in batches of 50 or so, the first ones ground will cool before they are scanned. The NIRS technique is based on spectra, and it is important to include samples of the same chemical composition that have been affected by factors such as growing location and industrial processing conditions (see Chapter 2, section 2.3).

The sources of variance in the *reference methods* should also be identified because the efficiency of those methods (see Chapter 3, section 3.2) are also affected by some of the factors that influence diffuse reflectance.

9.1.4 Step 5 – assembly of samples with variance

This is the most demanding and time-consuming, and an expensive part of the development of an NIRS calibration model due to the costs of laboratory reference analysis. Samples must be assembled to include all of the sources of variance that have been identified. In the case of many materials, moisture usually varies independently of other chemical constituents. It is difficult to assemble sufficient samples carrying the required range in moisture at the same time as the range in other constituents. Moisture ranges can be generated in wheat and barley and some other cereals by tempering, but grains such as maize, soybeans and oilseeds cannot be tempered so effectively, while it is impossible to induce ranges of moisture artificially in many materials. In terms of grains and forages, the best time to assemble samples for calibration development for moisture is at harvest time when the moisture range is usually at its extremes. Samples must be carefully protected from changes in the moisture content by storage in heavy plastic bags, such as Ziploc®, and scanned and analysed quickly to avoid development of moulds. The labile nature of moisture content is a serious source of error in development of calibration models for accurate prediction of moisture by NIR reflectance or NIT instruments.

Variance induced by the growing season is the most time-consuming factor in developing a stable calibration model in growing crops because it takes several years. Growing seasons appear to have a unique tendency to introduce spectral variance mainly due to changes in texture. The best way to correct for this is to monitor the first few (20–50) samples harvested in a new season, to detect whether the existing calibration can be used to test the new season's samples without bias. If bias does appear, the calibration should be strengthened by adding new samples early in the season. Addition of at least 30 samples of the new season with a range of composition, is recommended. After a few seasons (usually 5 or 6), the calibrations should be sufficiently stable to include any subsequent seasonal variability. The ANN calibrations for wheat protein and moisture have proved to be extremely stable to the seasonal effect. Because of environmental factors, such as temperature and humidity, seasonal variance is applicable to many agricultural and biological materials, including those of animal origin.

The stage of growth is an important factor in developing calibrations for forages. Lignification that begins to develop as the plant matures changes the texture of the material. Lignin is a complex polymer that cross-links with cellulose and adds strength to cell walls. This allows maturing plants to support the growing weight of seed development, and literally keeps the heads above ground. Lignification affects the particle characteristics of the ground sample that has been prepared for reference analysis and affects the precision. For animal feed mills, variance in the sources of bulk ingredients, such as grains and supplements should be included in samples assembled for calibration development. Flour mills recognise that the main source of variance to consider lies in their intake of wheat, especially if the mills are purchasing samples directly from farmers.

As an example, consider the spectra of two samples of the same wheat variety and genetic constitution when grown in the northern wheat-growing areas of Alberta and in southern Saskatchewan, Canada. The samples may be reported to contain the same protein and water contents, but the different growing conditions can affect the molecular arrangement in space and, as a consequence, the relationship between the spectra and the reported chemical composition that results from the calibration process. Variety (genetics), growing location and season, seed texture, composition, storage, cultivation practice, as well as soil terrain and fertility level are all factors that can affect the molecular arrangement and spectra of grains and plant materials, but not necessarily the chemical composition as measured by the reference method. Also, silages will be composed of a range of plant species and substances of different spectral composition, but can also be of similar composition as measured by chemical analysis.

For industrial-scale calibration development, replication of samples of agricultural materials that are reported to be of the same composition is strongly recommended. Differences in spectral composition, incurred by the differences in growing conditions that affect molecular and spectral composition, confer stability to the calibration. The application of chemometrics and pretreatments, such as derivatives, are based on the original absorbance data of the samples as scanned. Such differences in spectral characteristics of the same types of material of the same reported chemical composition, offer a possible explanation for the occurrence of what may be regarded as outliers.

When in use, NIR spectrophotometers regard their calibration equation as a constant, and apply the equation directly to any sample that they scan. Spectral precision is usually superior to that of laboratory analysis. If a significant variable, such as an important growing location, has not been included in the samples used in calibration development, the instrument will predict its composition using the calibration. The NIR technique is based totally on the spectra, and such samples would be predicted the way that the spectra appear to the instrument. If the results are not in accordance with the sample matrix assembled for development of the calibration equation they would be outliers, of which the operator would not be aware, because the results would be of the same order of magnitude as all of the other samples scanned. Provided that the samples assembled for calibration development have included all of the identifiable sources of variance for commodities such as grains and seeds, and even high moisture materials such as silages, the NIRS predicted results can be expected to be more reliable than those of the laboratories from which the reference results originate. This is partly a function of the spectral precision, and partly of the larger sample size as scanned by the spectrophotometer.

Assembling samples for calibration for applications where uniformity in the end product is the goal, is complicated due to the lack of variance. Flour mills can overcome this by using millstreams that differ in composition and particle characteristics. Industrial plants, such as feed mills, can generate calibration samples by creation of samples with a range in composition wide enough to provide a stable calibration. When calibrations are developed on these types of industrially prepared samples, fewer samples are required, but changes in composition, inclusion of different mill streams should be replicated at least twice, and preferably more frequently. The lower variance (SD) of the samples will affect the RPD, and Table 4.19 (see Chapter 4, section 4.1.13) is more applicable to evaluation of these calibrations. Samples for calibration of at-line instruments that are to be used to scan products or product mixes on moving belts can be

calibrated with static samples, and the calibration evaluated by subsequent removal and analysis of samples from the moving belt or mixer.

There has been some interest in development of calibrations for prediction of parameters, such as alpha-amylase (sprout damage) or mycotoxins in grains. The calibrations are believed to be feasible because of changes in the texture of grains induced by, in these cases, sprouting of the grains or penetration of the grains by fungal hyphae. The term 'double variance' is introduced to explain the rather poor efficiency of these calibrations. In a sample of wheat that is reported to have a DON content of 2.0 ppm, only about 3–4% of the kernels may be affected by the fungus (Fusarium graminearum). This means that in a sample of 100 g only about 100 (of 2860) kernels will visually appear to be affected. But these can each vary from a few ppm up to over 200 ppm in DON. At a visible damage level of only 3–4% only a few of these will actually get scanned by the instrument. This variance within the overall variance (double variance) is believed to interfere with the interaction between the instrument and the sample. The same principle applies to sprout damage in wheat or barley in that only a small percentage of the kernels may have visibly sprouted, but a higher proportion are ready to sprout, and the alpha-amylase content can vary widely among the kernels.

The most important factors affecting soils are the physical composition and moisture content. Soils vary in texture from very light sands and loams to heavy intractable clays, and this affects presentation of the sample to the instrument, as well as sample preparation for reference analysis. Drying and processing soils, manures, composts and similar materials for reference analysis cause irreversible changes in texture and other physico-chemical parameters. These materials should be scanned in their natural state as far as possible. Slurries such as liquid manures vary in total solids content, which changes their diffuse reflectance or transmittance characteristics (see Chapter 8, Figure 8.7). Liquid manures and milk both tend to sediment very quickly on standing and must be agitated thoroughly before scanning or reference analysis. Milk differs in that its fat globules sediment upwards. Drying manures and silages can cause losses of volatile constituents, such as ammonia and low-molecular-weight fatty acids. Continuous on-line application of NIRS to liquids is becoming increasingly exploited. It largely obviates the need for sampling, sample preparation and sample presentation, and overcomes errors due to sedimentation.

9.1.5 Step 6 – identification of a sample preparation system

The basic steps in sample preparation include thorough blending, sub-sampling, documentation, removal of foreign material, moisture reduction (drying) and size reduction by chopping or grinding where necessary, re-blending and storage until scanning or analysis. In the case of liquids, accurate dilution may be needed before reference analysis. For materials such as liquid manures, worts and mashes, NIRS can be carried out on the original material, whereas dilution, sub-sampling and perhaps filtering are necessary for practically all reference analysis. For most precise results, the moisture content should be determined before and after sample preparation. In fresh forages and fruits, liquid manures, worts and similar materials it is the data on total solids, rather than the moisture content, that are reported. The NIR instrument will scan the sample as-is, and will record spectral data on the as-is moisture basis. For materials, such as ground grains, meals and flours where the range in moisture content is usually relatively small, calibration models can be developed using reference results that have been reported on a constant moisture basis, such as 13.5%. Once a sample preparation system has been identified, it should be tested for reproducibility in both reference analysis and NIR spectral precision. Areas where errors might occur can be identified and steps taken to correct for them.

Excel files can be set up for documentation of all samples as they are received. The reference data can then be recorded in these files, which makes it easy to import reference data to NIR spectral files. Both spectral and reference data can be sorted before importing the reference data from Excel.

9.1.6 Step 7 – preparation of samples for scanning

Sample preparation has been discussed in Chapter 7 (section 7.5). Efficient blending before and after sample preparation is essential. Sample preparation is often regarded by staff as a dreary chore, and it is here that mistakes can occur frequently, due to boredom and carelessness. But efficient sample preparation is vital to the effectiveness of subsequent analysis by both NIRS and reference methods. Sample preparation should always be assigned to reliable hands.

9.1.7 Step 8 – Identify or develop a sample presentation system

Most NIR instruments are equipped with a cell for presentation of the sample to the instrument. This may be in the form of a hopper into which the sample is poured. For on-line applications, a sensor or sensors have to be positioned to receive the NIR signals from the sample. The positioning of the sensor must be

optimised to give the best resolution and precision of the signals. In the case of fresh materials, such as fruits, vegetables, forages and meat, the sample presentation system may be unique to the type of material, and development of the system may involve the method of size reduction (e.g. by chopping), preparation of the surface of the material, positioning of the sensor(s) and other features. Specialised sample presentation systems may have to be developed between the client and instrument company. The specialised engineering may incur significant extra expense, but again, this should be assessed by comparison to the improvement in the overall efficiency and expense of running the operation. It is important to determine the spectral precision of such sample presentation systems (see section 9.1.2).

9.1.8 Step 9 – repeatability file

A 'repeatability file' helps to stabilise calibration models for analysis of solid materials. The concept was introduced by Infrasoft International. The file is set up by selecting a small number of samples of the material to be scanned that represent a range in composition, particularly of moisture content, in agricultural and food materials. The samples are packed in cups, sealed and scanned. The spectra are stored in a repeatability file. The samples are then subjected to variations in temperature, over the range anticipated during day-to-day work, re-scanned, and added to the file, under the same sample name. Other factors may be identified that can affect the spectra, and the samples should also be exposed to variations in these conditions, and re-scanned. The reference data used in the repeatability file are the same as those of the original samples. The repeatability file is incorporated into development of the calibration model by the WinISI software. Repeatability files can be extended by adding more samples and/or operating conditions.

The efficiency of a calibration that has incorporated a repeatability file can be verified by using in the validation sample set 2 or 3 different samples of the same material that have been subjected to the same variations as the repeatability file in operating conditions. These should predict as accurately as all of the other samples in the validation set.

For operators that do not use WinISI, a repeatability file can be developed by selecting a small number of samples with a range in composition that have already been included in the calibration set. These are subjected to a series of differing temperatures and other conditions. The samples are re-scanned and included in the calibration file. Because some software recognises values of zero

as such, the original reference data of the samples should be inserted. Exposure of the samples to the different operating conditions will introduce spectral variance that will serve the same purpose as the repeatability file in WinISI. If the changes in operating conditions are likely to affect the moisture content, the samples will need re-analysis.

The main effect of the repeatability file is to introduce extra spectral variance at different levels of composition, and with other factors that can affect the spectral data of a sample. It is particularly useful when used in conjunction with a small sample set. When a calibration is to be developed for a new constituent or functional parameter, for which only a few (e.g. up to 20) samples are available with reference data, the model can be used to analyse new samples and the NIRS predicted results compared statistically to the new reference data. The SEP (standard error of prediction) will change as new samples are added, and will indicate when the model has become reliable for actual use.

9.1.9 Step 10 – scanning the samples

Scanning the samples is the simplest step. The manufacturers' manuals explain the methods for individual instruments. Careful and consistent filling of sample cells is critical to reproducible NIRS analysis and is part of the analytical precision. This is not a factor with instruments that use the hopper sample access system, provided that enough sample is added to enable all of the increments specified. If the sample size is slightly too low in some cases, small amounts of the sample can be recycled during scanning to prevent abortion of the scan. With instruments that use a sample cell that is scanned from below, the amount of sample added to the cell must give a layer thick enough to guarantee complete diffuse reflectance. Usually 3 cm gives sufficient depth, but with large grains such as maize, 5 cm will provide a safe margin.

Viewing the spectra is possible with scanning spectrophotometers and interferometers. This enables the operator to detect abnormalities in the spectra. For example, the shape of the spectra will be abnormal if the wrong sample cell has been selected in the software set up, or if the cell has not been properly filled.

Figure 9.2 shows deformities in spectra of a fibrous commodity in comparison to a normal spectrum. Varying degrees of abnormality can be seen in the region of 920–1040 nm. These were caused in this case by incomplete filling of the sample cell because the sample size was too small. A good calibration was obtained by avoiding the area of the anomaly. The large 'peak' at about 680 nm was caused by chlorophyll and was not an anomaly.

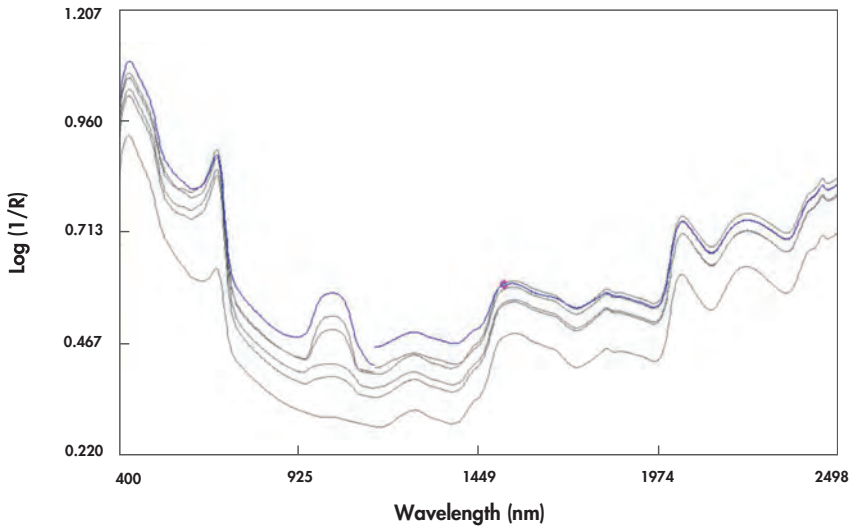


Figure 9.2 Abnormal and normal spectra of straw. Note the distortion around 1000 nm.

9.1.10 Step 11 – reference analysis

This step involves the performance of the reference analysis and addition of the results to the file containing the spectral data. Be sure to add the reference results to the right spectra. About 2% of NIRS outliers are caused by attempting to 'marry' the wrong partners.

The importance of accurate and precise reference analysis cannot be over-emphasised. In terms of day-to-day analysis where money is concerned, such as at grain delivery points, there are legal aspects that have to be aligned with the reliability of reference analysis. Contracts have to recognise the implications of the SEP and the consequences of error. Contracts should be based on the accuracy that is achievable as indicated by the 95% confidence intervals on the constituent, rather than the SEP. For example, a contract that demanded 'accuracy' within 0.1 or even 0.2% for something like protein content in grain would be unrealistic and would be open to a swarm of complaints from disgruntled clients. A SEP of 0.15% (in protein content) is excellent for either NIRS or reference testing, but only 68% of all samples will fall within the limits of ±0.15. The next 27% of samples will fall within limits of ±0.3, the next 3% within limits of ±0.45 and the remaining 2% outside of even these limits. Table 9.6 gives examples of the size of errors that can occur for different values of the SET (NIRS or reference methods).

Moisture content of the initial material before testing becomes even more important with fresh or high moisture material for two reasons. First, it is important to convert the analytical results from the dry matter stage normally

used for reference analysis, to the original moisture content of the material so that the calibration is based on the actual composition of the material as tested. The second reason is the effect of the change in moisture on the statistics of the calibration. For example, if the reference analysis reports a range in protein content of 12–17% but the moisture content of the fresh material was 66–70%, the actual protein content of the material as scanned would range from 4.2–5.8%, and the SD of the reference data would drop from 1.08 to 0.55. This would affect the efficiency as well as the statistics of the calibration.

Table 9.6 Implications of error in protein testing of wheat

NIRS result = 12.5%	SEP = 0.15%		SEP = 0.25%	
Limits of accuracy	**Maximum**	**Minimum**	**Maximum**	**Minimum**
68% of results	12.6*	12.4	12.8	12.2
95% of results	12.8	12.2	13.0	12.0
99% of results	12.9	12.1	13.1	11.9

*Results rounded to one decimal place at elevators

9.1.11 Step 12 – selecting samples for calibration

To obtain a rough evaluation of whether NIRS is applicable to the material and parameter to be measured, a calibration model based on a calibration set of about 40–50 samples with a validation set of 12 to 18 samples, or even just cross-validation, will provide a quick answer. In the early days of NIRS application, 40 samples were considered to be enough, mainly because of restrictions imposed by the costs of reference analysis, the time of computing and computer memory. Advances in memory and computing time (and experience) have removed the latter restrictions. The costs of reference analysis remain, but if the instrument is to be used in an industrial operation, the cost of reference testing are very small compared to the value of the testing to the operation. For testing a commodity that is not expected to show much annual or batch-to-batch processing change, fewer samples can be used, provided that all sources of spectral variance are well-represented.

The question 'how many samples do we need?' is frequently asked. The answer is the more the better as long as the samples accommodate all of the sources of variance. One guideline is to assemble 20–25 samples for each factor that you intend to use in development of the calibration. Of recent years, PLS (partial least squares) regression has become the method of choice for calibration model development. If 15 PLS factors are specified, at least 20 samples should be

assembled per factor, for a total of 300 samples. For MLR calibrations, 25 samples per wavelength are recommended, i.e. for a model of nine wavelength points this means assembly of at least 225 samples. The actual number of wavelengths used in the final model will be optimised, but a starting point of 9 with 25 samples per wavelength will provide a good sample set. Samples should always be selected bearing in mind the sources of variance. If eight growing locations are to be represented, samples from all of the locations must be assembled. These should also represent the range in composition anticipated. NIRS analysis derives solely from the spectra. Because of external influences, such as growing location and season, samples of agricultural materials (such as grains) that are reported as having the same composition by chemical analysis, may have different spectral composition – which will increase the total number of samples required. These extra spectral data must be included, which is why it is important to identify the possible sources of variance before starting to assemble samples for calibration model development.

Sample assembly is the biggest job in developing calibration models for NIRS analysis. When an operation is beginning to use NIRS for the first time, it is necessary to build up a database. It may not be possible to obtain 200–300 samples with the required variance all at once. So it is reasonable to begin with whatever samples are available with reference data, and add more samples as they arrive. Operators that seek to develop their own calibration models can start by assembling a database of samples with reference data even before acquiring an instrument. The samples should be carefully stored, ideally refrigerated but not frozen. When an instrument is acquired, a preliminary calibration can be developed using these stored samples (e.g. as few as 30 or so) and applied to every new sample as it arrives. The new samples would be tested by reference methods, and samples that have been added to the spectral and/or composition variance would be used to extend the calibration model, even for the more expensive reference methods, such as digestibility and metabolisable energy. As these are tested the database can be extended, until over a few years a comprehensive database has been built up. This method of establishing a calibration model is helped by using a repeatability file to add extra variance. A repeatability file is very useful when developing a calibration with a relatively small sample set because it adds spectral variance on the commodity without affecting the reference data (see section 9.1.8). The same repeatability file of the same material can be used in the development of calibrations for any number of constituents/parameters.

For development of calibrations that can reliably be transfered to a network of instruments all of the same make and model, spectral data from samples that have

been scanned on a group (at least 5, preferably more) of these instruments should be included. The extremely stable ANN calibrations supplied by some instrument companies incorporate over 70,000 spectra with reference data, recorded on a large number of instruments. There will be a range in the accuracy of the reference data. However, provided that this is unbiased, the inclusion of variance in both spectral and reference data adds a type of 'sample noise', which stabilises the calibration model (see Chapter 4, section 4.1.10).

The two main methods for sample selection are: a) *selection on the basis of reference tests*, and b) *selection on the basis of spectral characteristics*.

a) *Selection on the basis of reference tests* – this assumes that the operator has access to sufficient samples with reference data to generate the calibration. If the samples are distributed in normal (Gaussian) fashion with respect to composition, the results of prediction will tend to regress toward the mean (Figure 9.3). This phenomenon is called regression to the mean. It causes results at the high end to appear to be lower than they are, while results at the low end will appear to be higher.

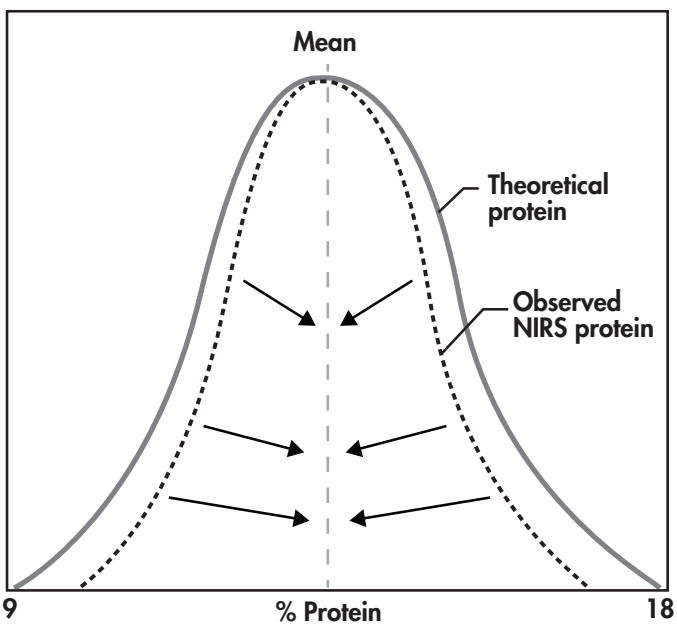


Figure 9.3 Regression to the mean.

The amount by which the NIRS result of a given sample will differ from the true result (actual reference) on prediction depends on the distance of the predicted value of the sample from the population mean, and the correlation coefficient (r)

between NIRS and reference data. The apparent deviations that can be expected can be calculated using the following formula:

$$\text{Apparent result } (x) = \bar{y} + (y - \bar{y}) \times r$$

Thus if, $y = 15.6$; $\bar{y} = 13.1$ and $r = 0.98$

Then,

$$\text{Apparent result } (x) = 13.1 + (15.6 - 13.1) \times 0.98 = 2.5 \times 0.98 = 13.1 + 2.45 = 15.55$$

Or if, $y = 15.6$; $\bar{y} = 13.1$ and $r = 0.909$

Then,

$$\text{Apparent result } (x) = 13.1 + (15.6 - 13.1) \times 0.909 = 2.5 \times 0.909 = 13.1 + 2.25 = 15.35$$

Or if, $y = 9.7$; $\bar{y} = 13.1$ and $r = 0.909$

Then,

$$\text{Apparent result } (x) = 13.1 + (9.7 - 13.1) \times 0.900 = -3.4 \times 0.900 = 13.1 - 3.06 = 10.04$$

Recall that for a given SEP, about 68% of all of the samples will be distributed within ± one SEP value of the mean result, and 95% will occur within ± twice the SEP value of the mean. The effect of regression to the mean is minimal if the correlation coefficient is high, but becomes sizeable at lower values of r. At stated values close to the mean the effect is also reduced, and an argument in favour of selecting samples on the basis of Gaussian distribution – because the majority of the population will differ from the mean by only ±1 SD-value, Gaussian selection will tend to favour the accuracy of the majority of samples to be tested. A set of samples with uniform distribution of components across the full range would, to a large extent, overcome regression to the mean, but assembly of such a sample set containing, e.g. 200 samples, would be a formidable task. For the purpose of day-to-day analytical work where large volumes of commercial samples are to be processed and the price depends on accurate analytical data, the Gaussian distribution is acceptable for sample assembly.

The best way to minimise the effect of regression toward the mean is to take every step to ensure the highest r^2-values in model development. Sample assembly becomes complicated if the objective is to develop calibrations with more or less uniform distribution of composition for more than one constituent or parameter, especially when the constituents/parameters are not correlated with one another, for example in the case of moisture and protein contents. When accuracy at extremes of reference data is important for some constituents, for example in plant breeding programmes, uniform distribution across the range of reference data is helpful, even if it is only achievable for those constituents.

When samples are selected on the basis of reference tests, all samples will have reference data. Recommended steps for setting up calibration and validation sample sets are:

1. Scan the samples.
2. Sort the samples on the basis of the reference data.
3. Select every 5th sample and save them as the validation set.
4. Save the remaining samples as the calibration sample set.

Sorting the samples on the basis of reference data in this way assures a similar distribution of spectra with associated reference data in both calibration and validation sample sets. If check samples or duplicates were scanned, they must be removed before selection of calibration and validation sample sets. Reference data for these can be set at 1.00 or 100, or some value that is distinct from the true reference data. It is then very easy to sort out the check samples. They should be stored in a separate file and can subsequently be predicted to determine the precision (standard error) of the NIRS test. The check samples and duplicates should then be purged from the main sample files before development of the calibration model.

If cross-validation will be used to evaluate the model (see Chapter 5, section 5.2.9), there is no need to sort the samples because all of them will be used in calibration and validation, but it is a good idea to sort the samples by composition, and set up a validation set anyway. Use the calibration set for model development and optimisation using cross-validation, and test it by predicting the validation set.

b) *Selection on the basis of spectral characteristics* – this concept was originally introduced by Technicon (later Bran+Luebbe) as a means of reducing the amount of expensive reference analysis required by selecting samples for calibration with maximum variance in spectral data. The technique has been effectively refined by InfraSoft International (ISI, Port Matilda, PA), and their WinISI software enables user-friendly selection of samples for reference analysis on the basis of spectral characteristics before any analytical work is done. The ISI principle is: if two or more samples are very similar spectrally, only one of them is needed. The software compares the samples on the basis of 'scores', that are developed from Mahalanobis distances, and selects those that offer the most diverse grouping of spectra. Over-simplified, the Mahalanobis distance of a spectrum/wavelength point is the distance in spectral data of a that spectrum from the mean of the population and from the other samples.

Samples can be selected on the basis of log 1/R data as recorded, or on the basis of pretreated (derivatised) spectral data. The steps are detailed in the WinISI

software materials (Infrasoft International, Port Matilda, PA, USA) and are briefly:

1. Scan the samples
2. Develop scores
3. Use the WinISI software option to select samples for the calibration set
4. Use this set for model development and cross-validation
5. Save the remaining samples (with scores) in a separate file and repeat the selection
6. Save this second file as a separate validation set
7. Send samples represented in both files for reference analysis

Steps 5 and 6 are optional, and unnecessary if cross-validation is to be used for evaluation throughout.

For setting up calibration and validation sample sets (or test set), a useful guideline is to use 80% of the samples for calibration and 20% for validation (see Chapter 5, section 5.2.4). A downside of the spectral selection system is that a large number of samples have to be assembled and scanned anyway to enable the spectral selection.

The WinISI system describes Global H (GH) and Neighbourhood H (NH) characteristics for all samples, which are based on Mahalanobis distances. The GH value is the distance in space between the sample and the mean of all of the samples. The NH value of a sample is the distance in space between any sample and its nearest neighbour. WinISI recommends that samples with GH values of above 3.0 should not be included in calibration development. Such samples are usually present as a small percentage of the population, but will probably persist in all samples of the same material for which calibrations are to be developed. Their inclusion in the calibration development will have a possible adverse effect on the calibration statistics, but their inclusion is also likely to improve the stability of the calibration to future analysis.

9.1.12 Steps 13 and 14 – development and optimisation of the calibration

Model development can be done using MLR or PLS regression. Instruments that use discrete filters usually use MLR for development of calibration models. A number of instrument companies have developed software specific for the computation. Software packages such as WinISI and VISION for use with the Foss/NIRSystems scanning instruments, and NIRCal for BÜCHI FT-NIR instruments, have both MLR and PLS regression options. OPUS for use with the Bruker instruments uses only PLS regression. These packages are dedicated to the instruments and

have the big advantage that the instrument can use the calibration models directly.

WinISI software includes the 'Local' option. This is a PLS-based system best suited for application to large databases. It has been shown to be superior to standard PLS and Modified PLS options. The principle of 'Local' is that it scans a large database (preferably several hundreds, or even thousands of spectra) and develops a 'model' (possibly best described as an 'attitude') based on all of the spectral and reference data. It then predicts any new sample on the basis of the spectral characteristics of a specified number of spectra in the database that most closely match the new sample.

For large-scale testing, including screening during breeding programmes, grain-handling operations and day-to-day routine analysis, which involves the analysis of large numbers of samples over long periods of time, the 'database' and 'Local' approach is likely to become the way of the future for analysis of the more frequently tested constituents. Any organisation involved in large-scale analysis of this type can and should accumulate a database of spectra with reference data. The 'Local' option of WinISI can be updated to include new constituents or parameters as the reference data for them becomes available. Operations that contemplate using NIRS in the future can start assembling samples to build a database before they buy the instrument. The samples would be tested by all reference methods and stored until they can be scanned. This is a useful starting base for development of stable calibrations, especially if used in conjunction with a repeatability file.

For operators that use The Unscrambler®, Grams®, Pirouette® or MATLAB® software, the spectral data have to be transferred from the instrument software to the preferred software and the calibration models developed – after validation the final calibration models are then transferred back to the instrument operating software. Software companies provide manuals for their operation, as well as background information on NIR technology.

For instruments that use discrete filters, log 1/R is the only form in which spectral data can be processed, and MLR is the method usually used for calibration development. The number of filters selected can be optimised by recording models for several filter combinations and tested by predicting a validation sample set. Operators of scanning spectrophotometers that use MLR for computing the calibration can select up to 9 wavelengths, but the number selected using MLR should be optimised by validation.

PLS regression uses all available wavelengths, but the most stable equations often do not need the full wavelength range from 400–2498 nm. Optimising the wavelength range usually leads to improved validation statistics and more reliable calibration models. The calibration equation is developed using different wavelength ranges and the same mathematical pretreatment of the spectra data. Table 9.7 shows results of optimising the wavelength range for prediction of protein content in whole wheat.

Table 9.7 Influence of wavelength range on prediction of protein content (%) in whole wheat

Wavelength range (nm)	SEP (%)*	Bias	RPD
408–1092	0.353	-0.136	4.4
408–2492	0.293	-0.164	5.4
658–1792	0.233	0.120	6.8
708–1092	0.244	0.096	6.4
908–1692	0.198	0.169	7.9
1108–1408	0.195	0.088	8.1
1108–2492	0.227	0.074	6.9

*Mathematical pretreatment was log 1/R, smoothed with segment of 4 wavelength points (WinISI terminology = 0 0 4 1)

The mathematical pretreatment of the spectral data was log 1/R using WinISI software and smoothing by 4 wavelength points. Four of the wavelength ranges are used in commercial instruments. The 658–1792 nm wavelenth range has been found to be very useful in some applications. The 708–1092 nm range covers the low wavelength range, but eliminates most of the noisy area below 700 nm. Modern computers and software have combined to make optimisation a simple if rather repetitious process, but all of the above wavelength ranges, and any others that are of potential interest, can be tested in a few minutes.

Once the most suitable wavelength range has been identified using, e.g. log 1/R data, the exercise should be repeated using mathematically pretreated spectral data. Pretreatment of the spectral data often improves the validation statistics. Most monochromator instruments record spectral data at intervals of 2 nm. Diode array instruments record at intervals of about 5 to 6 nm. The most successful pretreatments are smoothing of the log 1/R signal, and developing the 1st or 2nd derivative of the log 1/R signal, usually on already smoothed data. The dimensions of the derivative has come to be referred to as the ‘gap’ and the

degree of smoothing the 'segment' (see Chapter 5, section 5.2.5). Table 9.8 shows the influence of optimising the mathematical pretreatment of spectral data for the prediction of protein content in whole wheat. The samples were the same as those used in compiling Table 9.7, and illustrate the value of optimising both wavelength range and mathematical pretreatment.

When working with a constant type of material, e.g. whole barley or fresh forages, a pattern will emerge, and one or two of the wavelength ranges will appear consistently as the most suitable. Thereafter, there will be no need to test all of the other ranges. With a substance such as soil, the wide-ranging variance in physical characteristics complicates calibration development, and it may be necessary to test several wavelength ranges and mathematical pretreatments for these types of materials from different sources. Note the improvement in the statistics when the wavelengths between 400–700 nm are avoided. Also note that within a wavelength range, the different mathematical pretreatments had relatively little effect. Changes in wavelength range usually have more effect than mathematical pretreatment.

Table 9.8 Influence of mathematical pretreatment of spectral data on prediction of protein content (%) in whole wheat for selected wavelength ranges

Derivative	Gap	Segment	WinISI notation	SEP	Bias	RPD
408–1092 nm						
None	0	4	0 0 4 1	0.353	-0.136	4.4
1st der*	4	4	1 4 4 1	0.298	-0.176	5.3
1st der	4	8	1 4 8 1	0.296	-0.155	5.3
2nd der*	4	4	2 4 4 1	0.352	-0.367	4.3
2nd der	8	8	2 8 8 1	0.280	-0.052	5.6
708–1092 nm						
None	0	4	0 0 4 1	0.244	0.093	6.4
1st der*	4	4	1 4 4 1	0.234	0.073	6.7
1st der	4	8	1 4 8 1	0.248	0.101	6.3
2nd der*	4	4	2 4 4 1	0.253	0.075	6.2
2nd der	8	8	2 8 8 1	0.277	0.063	5.7
1108–2492 nm						
None	0	4	0 0 4 1	0.228	0.068	6.9

Table 9.8 Influence of mathematical pretreatment of spectral data on prediction of protein content (%) in whole wheat for selected wavelength ranges (continued)

Derivative	Gap	Segment	WinISI notation	SEP	Bias	RPD
1st der*	4	4	1 4 4 1	0.189	0.084	8.3
1st der	4	8	1 4 8 1	0.202	0.123	7.8
2nd der*	4	4	2 4 4 1	0.200	0.108	7.9
2nd der	8	8	2 8 8 1	0.202	0.092	7.8
908–1692 nm						
None	0	4	0 0 4 1	0.199	0.169	7.9
1st der*	4	4	1 4 4 1	0.183	0.088	8.6
1st der	4	8	1 4 8 1	0.185	0.093	8.5
2nd der*	4	4	2 4 4 1	0.190	0.103	8.3
2nd der	8	8	2 8 8 1	0.184	0.096	8.5
656–1792 nm						
None	0	4	0 0 4 1	0.233	0.118	6.8
1st der*	4	4	1 4 4 1	0.179	0.081	8.8
1st der	4	8	1 4 8 1	0.183	0.104	8.6
2nd der*	4	4	2 4 4 1	0.187	0.091	8.4
2nd der	8	8	2 8 8 1	0.182	0.117	8.6

*1st der = first derivative of the log 1/R spectral data; 2nd der = second derivative of the log 1/R spectral data; *WinISI notation: 0 0 4 1 = log 1/R, smoothed by 4 points; 1 4 4 1 = 1st derivative with gap and segment of 4 points; 1 4 8 1 = 1st derivative with gap of 4 and segment 8 points; 2 8 8 1 = 2nd derivative with gap and segment of 8 points

Table 9.9 shows the influence of optimising both wavelength range and mathematical pretreatment on prediction of NDF (neutral detergent fibre) content in ground barley. Note that the optimum mathematical pretreatment of the log 1/R spectra data differed depending on the wavelength range. All calibrations were developed from samples scanned in reflectance mode. The RPD values for wheat protein in Table 9.8 were generally better than were those for prediction of barley NDF. This is related to the standard error of the respective tests. NDF is subject to more sources of error than the combustion (Dumas) method used for determination of protein content. Notice that wavelength range generally again has more influence on calibration efficiency than mathematical pretreatment.

Table 9.9 Influence of wavelength range and mathematical pretreatment on prediction of NDF content of barley

Wavelength range (nm)	Mathematical pretreatment*	SEP	Bias	RPD
408–1092	1 4 4 1	0.708	0.188	4.0
408–2492	0 0 4 1	0.574	0.079	5.0
658–1792	1 4 4 1	0.509	0.176	5.6
708–1092	0 0 4 1	1.152	-0.081	2.5
908–1692	1 4 4 1	0.597	0.209	4.8
1108–1408	2 8 8 1	0.643	0.012	3.3
1108–2492	1 4 4 1	0.572	0.200	5.0
1208–2392	1 4 8 1	0.584	0.122	5.0

*WinISI notation: 0 0 4 1 = log 1/R smoothed by 4 points; 1 4 4 1 = 1st derivative with gap and segment of 4 points; 1 4 8 1 = 1st derivative with gap of 4 and segment of 8 points; 2 8 8 1 = 2nd derivative with gap and segment of 8 points

Some NIRS software, such as Bruker OPUS and BÜCHI NIRCal, include an optimising option. The calibration and validation sample sets are identified to the system, and the optimum wavelength range and mathematical pretreatments are both sought by the software. The calculation takes rather longer than 'normal' calibration development, but saves the time of researching different wavelength ranges and mathematical pretreatments.

9.1.13 Step 15 – entering the model into the instrument

This is carried out simply by software steps. Where the models have been developed by other than dedicated software the model has to be transferred from The Unscrambler®, Grams® or other software into the instrument software. The instrument operating manuals describe the steps for importing the model. Optimisation of wavelength and mathematical pretreatments should be carried out in all of these comprehensive software packages before model transfer.

9.1.14 Step16 – verify accuracy and reproducibility of NIRS-predicted results (evaluation)

There are three main methods for evaluation of calibration models. These are cross-validation, test-set validation and prediction of completely unknown samples. These are discussed in depth in Chapter 5 (see section 5.2.9). One simple message is that duplicate testing only takes a very short time by NIRS testing, and adds to the reproducibility and overall integrity of the reported result.

9.1.15 *Step 17 – verify accuracy by analysis of new samples of known composition*

When the model has been satisfactorily developed, the overall picture is completed by verification of the reproducibility of the NIRS analysis using the new model. The ultimate test of a calibration is to test new, 'fresh' (completely unknown) samples. These are 'real world' samples of the type that will form the future workload. A check sample or samples should also be tested periodically to monitor the precision. After analysis of a number of new samples using the calibration model, some of them should be sent for analysis by the reference method to verify the accuracy. The full statistical treatment should include computation of the regression (slope) as well as the SEP. If biases occur, the adjustment should be made and then further samples analysed to verify that the accuracy is acceptable.

9.1.16 *Steps 18 and 19 – carry out slope and bias adjustments as necessary and re-analyse the same samples to ensure that the bias has been removed*

If biases occur the adjustment should be made and then further samples analysed to verify that the accuracy and precision are acceptable. Slope adjustments are generally not recommended (see Chapter 4, section 4.1.7).

9.1.17 *Step 20 – supplement calibration model with future samples analysed for monitoring*

This is not actually part of the calibration process, but it is a part of on-going calibration maintenance. From time to time, new samples can be selected to extend the calibration model. This is advised whenever samples arrive from a new source, such as a new location or growing season, or when processing conditions change, such as in an industrial plant. As part of the continuing monitoring process, during day-to-day analysis of incoming samples, some samples can be identified as adding to the range of variance that has been incorporated into the calibration. They may have been identified as 'outliers' by the instrument based on their spectral characteristics.

Provided that the authenticity of these samples has been established as actually adding new and useful variance and that more samples of the same type can be obtained, the samples can be stored and used to extend the existing calibration model. This is if these samples are not likely to exert sufficient leverage to change the whole model radically. This is done by adding the samples to the database used for calibration, repeating the calibration process and entering the new calibration model into the instrument. Some of the new samples should also be compiled as

a test set to make sure that the new samples are predicted accurately after the extension of the model. Most typically samples representing new variance will appear early in new crop years, but they can also appear from different growing locations, different deliveries of ingredients, changes in processing conditions and other sources.

9.1.18 Step 21 – calibration transferability

If the operation is using more than a single instrument of the same make and model, the calibration model can be transferred to all of the instruments. Ideally, the instruments should be networked so that calibration transfer and performance monitoring can be controlled from a single computer (see section 9.2). Networking can be carried out using manufacturer-supplied software. The identification system for networked instruments can be based on the instrument serial number.

Calibration transferability depends on the type of instrument. Calibration transfer is generally more successful with monochromator- and interferometer-driven instruments than with other types of instrument. This is mainly because although minute differences do exist among individual instruments of these types, differences that occur among discrete filters (of the same nominal wavelength), diode arrays and CCD (charge-couple device) arrays, tend to be greater. To transfer calibrations among diode array instruments assembling spectral data from at least 5 instruments is recommended. That is, the training sample set would be scanned on all five instruments and the five data sets combined in the calibration development. The resulting calibration should be transferable to other instruments within and outside the group of five. The ability to transfer calibrations among instruments is an important criterion in evaluation of a new instrument and a factor in selecting an instrument for purchase.

9.2 Networking

Large-scale testing usually involves testing at more than one location. In the case of NIRS testing this means more than one instrument. The most efficient way of achieving this is by setting up a network of instruments. The instruments should all be of the same type and model and should all use the same calibrations. The networked instruments can all be controlled at a central laboratory from a single computer using the 'master-slave' concept.

When an organisation is operating several instruments, they can all be operated using a networking system. The instruments can all be monitored and controlled from a single personal or main-frame computer at a central laboratory or office.

Successful application of NIRS networking depends on the reliability of ten factors:

1. Reference or 'check' samples;
2. a system for distribution and use of check samples;
3. a dependable laboratory for reference testing;
4. a set of reliable instruments with reliable calibrations;
5. a central computer (this can be a laptop);
6. software for monitoring purposes and applying corrections to instruments;
7. software for development of calibrations;
8. software for operation of the network;
9. the person in charge of the networking operation; and
10. education and training of staff in all aspects of the operation.

The principle of networking is based on central control of the accuracy and monitoring of the precision (reproducibility) of the on-site instrument(s), all of which carry the same calibrations (Williams & Antoniszyn, 2004). The on-site operator usually has no access to the system for adjusting the instruments. Most large-scale networking operations have been established by commercial and government establishments that are involved with large-scale grain-handling. These have been developed in-house. Although there is an extensive bibliography on NIR technology, there is very little published information on NIRS network development. Table 9.10 provides steps to be followed when establishing an NIRS network system.

Networking is a way of broadening and getting the maximum benefits from the application of NIR technology. Instruments of the same type and model can be standardised and networked so that they all use the same calibrations and operate with the same accuracy and precision. Calibrations developed by one centre can be shared with other locations that use the same type and model of instruments. Another aspect of networking is that instruments can be located at strategic points in an industrial plant and by networking to one computer all aspects of the process can be regulated and controlled, including monitoring of the composition of raw materials, the efficiency of blending during processing and the composition of the final products. All adjustments are made on the basis of check sample results as tested by the master instrument. The master instrument is the only one that is monitored on the basis of reference laboratory results.

Table 9.10 Steps in establishing an NIRS network system

Step 1	Determine whether an NIRS network will be of economic and/or practical benefit to the operation
Step 2	Identify the items involved, and prepare a budget for setting up the network
Step 3	Identify control centre
Step 4	Identify method for networking (e.g. modem, e-mail, wireless communication)
Step 5	Identify a person to accept responsibility for running, monitoring, and controlling the entire operation
Step 6	Identify locations for instruments (at remote locations or at strategic locations within a processing plant)
Step 7	Install dedicated telephone line, or Local Area Network (LAN) system if modems are to be used with telephones.
Step 8	Identify reference methods and reference laboratory
Step 9	Identify the instruments to be used
Step 10	Identify system for monitoring (e.g. frequency of check sample testing, system of sample distribution, use of on-site testing)
Step 11	Identify source of check samples
Step 12	Assemble and prepare check samples
Step 13	Identify system for distribution of check samples
Step 14	Identify method for correction of instrument deviations
Step 15	Identify the Master instrument
Step 16	Assure that all satellite instruments or sensing heads are functioning properly
Step 17	Standardise all satellite instruments to master instrument
Step 18	Develop or acquire software for operation of network
Step 19	Train staff in use of network software
Step 20	Test network using limited number of locations
Step 21	Develop and apply corrective measures to system, as necessary
Step 22	Verify that corrective measures have been successful
Step 23	Extend the network to all locations
Step 24	Test networked instruments at all locations.
Step 25	Prepare a detailed step-wise manual for operating the system for use in training new staff, and ensuring that all locations are using exactly the same procedures

Sometimes a laboratory may wish to network different instrument models from the same company. The network will work reasonably well, but will use the wavelength range of the instrument model with the smallest range. For example, if two instruments that operate over a wavelength range of 400–2500 nm are networked with a third appliance that uses the range of 900–1700 nm (the diode-array instrument range), the network will operate on the wavelength range of 900–1700 nm. The wavelengths below 1100 and above 1700 nm will not be used by the first two instruments. The future challenge in networking will be to develop networking software that will use instruments of different models and companies. Software has also been developed for networking monochromator-driven with diode array-driven instruments. The same restrictions apply with regard to wavelength accessibility.

9.3 Last words on calibration development

Whether the calibration model is developed by the operator or the instrument company, the basic steps described in this section are the same. This chapter has been included for users who either have to develop calibration models, or wish to 'do their own thing'. It is included also to serve the needs of research stations and laboratories worldwide, including those in developing countries, where large-scale networking, and easy access to the parent instrument company and its expertise is not readily available. Calibrations will always be needed for NIRS analysis of fresh materials, such as fresh forages, fruits, vegetables, meat, fish, manures, and other high moisture materials. Calibrations will also contimue to be needed when a new functional parameter calls for rapid analysis. Files of spectra should always be retained and backed up. An external hard drive is an excellent and inexpensive method for avoiding the possible frustration of losing data. Many instruments can now be purchased with factory-developed calibrations. The essential requirements for calibration development are samples that represent all of the anticipated variance with reliable reference data. Instrument companies develop calibrations as described at the beginning of this chapter. The ANN calibrations offered by some companies are based on these types of spectra. Calibration models developed by instrument companies are marketed, and the prices vary depending on the complexity of the model and the parameter for which it has been developed to measure.

Reference

Williams, P. & Antoniszyn, J. 2004. Instrument Networking: Design and Implementation. In: J. Workman, C. Roberts and J.B. Reeves III (eds). *Near-infrared Spectroscopy in Agriculture*. Madison, Wisconsin: The American Society of Agronomy, Inc., The Crop Science Society of America, Inc., The Soil Science Society of America. 49–73.

10

Interpretation of calibration evaluation

In his classic biography of Michelangelo Buonarotti, *The Agony and the Ecstasy*, author Irvine Stone mentions that Michelangelo noted in his diaries that he could visualise statues, including his exquisite statue of David, inside the blocks of stone, and that all he had to do was to bring them out. An NIR spectrum can be considered in the same way. All of the information relating to the sample is contained in the spectrum. What we as operators have to do is to get the instrument to translate the spectral data into the information we need. Modern NIRS data analysis software contains many options, the chisels that can help us to bring it out of the stone.

This chapter is included mainly for the benefit of those who want to learn more about NIR technology than just its application. Most operators are content with scanning each sample once, or doing a single test on the sample by the reference method. The logic here is that in day-to-day analysis there is usually only time for a single NIRS test, and doing any more than a single test by reference methods adds to the costs and time of developing the calibration model. But again, these costs must be considered in the context of how much money and possible liability will be saved by using a more reliable NIRS calibration, and a more reliable way

of doing the test. Duplicate scans can improve the reliability of NIRS testing, and only take an extra minute or two.

10.1 Scope of calibration evaluation

Interpretation of calibration evaluation means finding out why the calibration worked, or why it did not work. The first step in calibration evaluation is to look at the statistics (see Chapter 4). The coefficient of determination (r^2), bias, slope and RPD figures are the simplest way of evaluating the calibration model quickly. But before doing any calibration work the error of both reference and NIRS testing should be determined (see Chapter 9, sections 9.1.1 & 9.1.2).

Most data analysis software packages offer systems for identifying outliers based on spectral characteristics. WinISI identifies them on the basis of their global 'H' (GH) values, and gives the option of removing them from the calibration sample set. GH values are based on Mahalanobis distances (see Chapter 9, section 9.1.11). They show the distance in spectral data by which individual samples differ from that of the population mean at each individual wavelength point. While GH values above 3.0 are often associated with large errors in prediction, it is possible to get a large error at GH values of even less than 1.0. Neighbourhood H (NH) values indicate the distance by which the spectral data of an individual sample differ from those of its nearest neighbour. Provided that there are no obvious oultiers caused by, e.g. inaccurate reference values, the elimination of samples with GH values of higher than 3.0 will improve the statistics of the calibration, but not neceassrily improve the integrity of the calibration. There are no outliers in the world of industrial application, and the inclusion of samples with high GH values should add a degree of stability to the calibration.

Outliers can also be identified at the time of evaluation by plotting the reference or actual against the NIRS predicted results. If a test set is used and the results are not acceptable even though no outliers are obvious, plot the reference against predicted data for the calibration set – one or more significant outliers may have been included in that sample set. These will affect the SEC and SEC), as well as changing the slope. An outlier can bias the entire model towards itself. Outliers should be investigated and if they are believed to be chemical outliers, the reference result should be verified by re-testing. If the outliers are spectral outliers, their origin should be determined and further samples from the same source added to the calibration (and validation) sample set before making a final evaluation of the calibration.

If cross-validation is preferred, the reference versus predicted results can still be plotted, and possible outliers identified and dealt with as for a test set. The number of outliers is usually quite small. If, on plotting reference against predicted results, it appears that there are a lot of outliers in a calibration (10% or more), it is unlikely that all of these are outliers. The most likely source of the excessive variance is in the reference data – again, the importance of knowing the error of the reference testing cannot be over-emphasised. 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, NIRS may be not applicable to that area of work.

10.2 Accuracy and bias

The efficiency of an NIR instrument for determination of a particular parameter is a function of the reference analysis, the calibration, the material to be tested and the actual instrument error. Once the calibration is developed and validated, accuracy can be changed to conform to what is considered to be the *true* result by changing the bias, but for practical purposes the results of NIRS will always be aligned with those of the reference laboratory, and the results as reported by that laboratory have to be accepted as the accuracy. Remember that neither the SEP nor SECV are indications of accuracy by themselves. It is possible to achieve excellent values for the SEP but with a significant bias (little or no information on bias is given by cross-validation). In reporting the results of an NIRS calibration evaluation, the SEP should always be accompanied by the bias. The on-going reliability with which this *true* result can be achieved is the precision (reproducibility) of the testing.

Precision is partly a function of the instrument, partly a function of the natural chemical and physical heterogeneity of the material being analysed, partly a function of the calibration, and partly a function of the operator in sample preparation and presenting the sample to the instrument. Accuracy is only established once (at the time of calibration). Thereafter it is maintained, as necessary, through monitoring. Precision affects every result because it happens with every test. For the most part, the instrument cannot be blamed for poor precision. While the importance of a high degree of precision is emphasised, the accuracy cannot be overlooked – otherwise all results will be reported precisely wrong! Bias can become costly, and every effort should be made to minimise and preferably to eliminate it. Small biases of for example 0.1% protein will happen. As long as they fluctuate between small levels, such as ±0.1 in protein content, they can be tolerated, but if they persist they should be corrected.

10.3 Graphical distribution of samples during calibration

Software such as WinISI and The Unscrambler® each include a system for viewing samples during calibration development. Figure 10.1 shows a 3-dimensional (3-D) WinISI PCA (principal component analysis) scores plot illustrating the distribution of samples used for calibration.

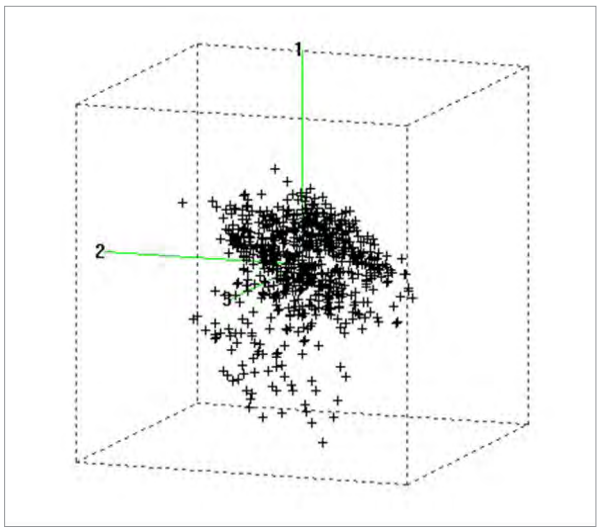


Figure 10.1 Distribution of samples during calibration (3-D WinISI plot).

Figure 10.2 is a 3-D plot (PC1 vs. PC2 vs. PC3), viewed from the right side of the above 3-D plot. In both figures trends are apparent, and show up rather more clearly in Figure 10.2. The samples were drawn from western Canadian plant breeders' material from 1998–2005. The main trends reflect different growing seasons in both figures. Samples at the bottom of Figure 10.2 are from the soft wheat class, and samples at the top right are of the western extra strong class, with much harder kernel texture. Right to left trends reflect different growing seasons. Samples to the right of the figure are from the 2005 season, and the earlier seasons are to the left.

Figure 10.3 shows the same type of 3-D plot taken only from the 2001 growing season. Here the different 'clusters' of scores reflect the different wheat classes. The cluster at the bottom of the figure represent the soft white spring class. The cluster on the extreme left are the CWES (Canada Western Extra Strong) class, the samples to the upper right include the CWRS (Canada Western Red Spring) and Central Prairie hard red winter class, and those in the center are of the western hard red winter class.

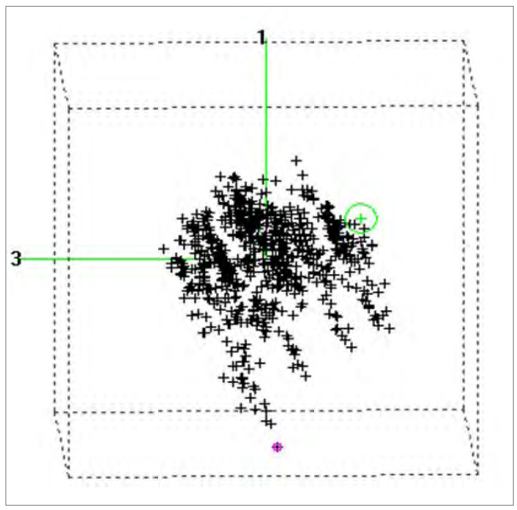


Figure 10.2 Distribution of samples during calibration for whole wheat PSI (right view, WinISI plot).

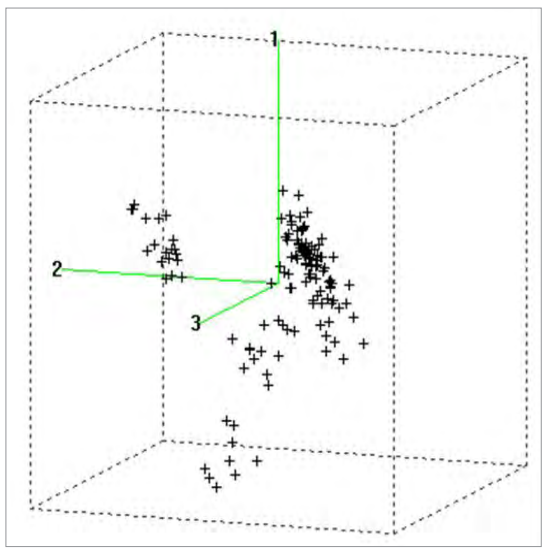


Figure 10.3 Scores plot of 2001 wheat samples, showing distribution of wheat classes (WinISI)

Figures 10.1 to 10.3 illustrate the influence of growing season and wheat class on the spectral characteristics of the wheat. All growing locations and seven seasons are included in the 875 samples of Figures 10.1 and 10.2.

Figure 10.4 shows a different system of studying what has happened during the development of a calibration model. The figure is taken from a calibration developed using The Unscrambler®. The figure shows a PCA scores plot of the first

principal component against the second (PC1 vs. PC2). The 'scores' are the position of the samples along the principal components (PCs,) and can be used to show similarities or differences among samples. The first two PCs usually explain most of the variance in most applications.

The display of scores in 2-dimensional (2-D) plots can show how the samples are distributed in the calibration and validation sample sets. Obvious trends in the distribution of samples can be traced by focusing on individual samples in the scores plots to identify the way in which the sources of variance in the samples have been interpreted by the analysis, and the way in which groups of samples interact with each other in the total population. Although trends may be apparent, in the world of practical application often several sources of variance will be involved in a population. It is useful to see the way in which various combinations of wavelength ranges and mathematical pretreatments resolve these to produce reliable calibration models.

The Unscrambler® offers several features that are useful in understanding what has happened during development of a calibration model. Figure 10.4 shows the distribution of samples by sample number from three classes of whole wheat, during development of a calibration for protein content. The positions of the points show that the samples were distributed according to the three classes. This plot usually appears in the top left panel of The Unscrambler® display after computing a calibration model.

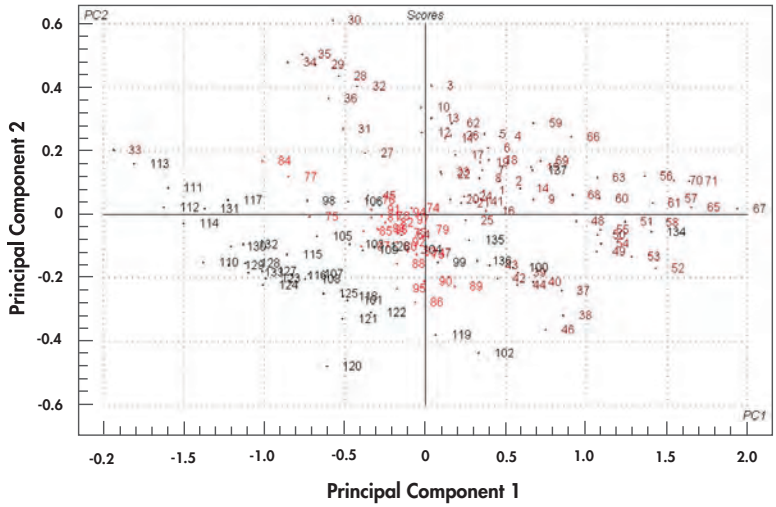


Figure 10.4 PCA scores plot (PC1 vs. PC2) of samples of three different wheat classes used in development of a calibration for protein content with The Unscrambler®.

Figure 10.5 shows the same scores, identified by category variables, in this case wheat class. A category variable is an option in The Unscrambler® that enables classification on the basis of the selected parameter. Category variables explain the origins of the trends. Note that it was possible to separate the samples from the two different (Central and Western) hard red winter wheat breeding programmes.

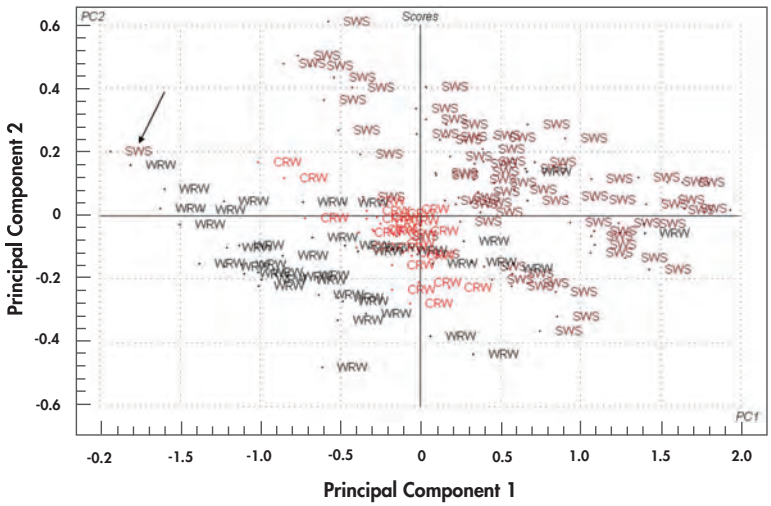


Figure 10.5 The same PCA scores plot as in Figure 10.4, labelled according to the category variable 'wheat class' (SWS = soft white spring; (CRW = Central red spring; WRW = Western red spring). One SWS sample appeared in the same group as WRW.

Notice also that one soft white spring (SWS) sample appeared in the same group as the western red winter (WRW) class. The breeder had included one hard white wheat line for one growing season, and the data analysis software was able to identify it as a hard wheat.

Kernel texture has a strong influence on spectral characteristics. Category variables can be used to identify spectra by their position and can be developed on the basis of features such as growing seasons, wheat class, growing location, forage composition, stages of growth, soil type and any other characteristics in any material. The data in Figure 10.5 represented three growing seasons and three classes of wheat. The main variables were growing season and kernel texture. The two winter classes were red wheat and the spring wheat was white. A second important difference was that the white wheat was a soft class, whereas both red winter wheats were harder in kernel texture. The tendencies were for growing season to increase from left to right, and kernel texture, in terms of particle size index, to move vertically.

10.4 Spectral basis of calibrations

10.4.1 PLS factors (loadings or weights)

A different system for interpretation of calibration results involves plotting the weights developed during computation of PLS (partial least squares) regression equations. This is done by viewing the PLS weights when the PLS regression is complete. The process of PLS regression is based on identification of the major sources of the covariance between the spectral data and the reference data. PCA searches for the largest variance in X (spectral data). PLS regression searches for the largest covariance (or correlation if X is autoscaled, i.e. centred and normalised) between the spectral and reference data. The vectors of PLS scores are also called 'latent variables'. The most important variables can be visualised with the weight (loadings) similar to PCA. PCA defines the components of variance in a system. Principal components are linear functions of the absorbances in the system, and contain all of the information in the spectral data. In NIRS work, PCA is carried out on the X-variables (the spectral data) to determine the distribution of the variance. The principal components are then regressed against the Y-variables (reference data), and used in prediction. This is PCR.

At any wavelength point there are several absorbers all with different degrees of intensity, and the highest correlation may not be associated solely with a 'classical' wavelength. The first principal component is the result of interaction among the 'mishmash' of absorbers, including combination bands and overtones, at that wavelength point. In PLS regression, the Y-variables are also used in estimating the latent variables (see Chapter 5, section 5.2.8) in order to confirm that the major components of the variance are the most suitable for prediction of the Y-variables. In theory, PLS regression should give results that are superior to PCR, but this is not necessarily true. If PLS proves not to be superior to PCR for a given exercise this could indicate that the reference analysis (the Y-data) has not improved the relationship between the principal components and the reference data, or that the reference data itself may not be reliable. This is yet another reason to ensure the precision of the reference method in NIRS application work.

The concept of PCA is easy to illustrate with the first two PCs, which can be graphically represented in the form of the x- and y-axis of a correlation scatter plot. The existence of a third PC is not difficult to imagine, as being the z-axis at right angles to the x and y of the scatter plot. But it is not so easy to visualise further PCs. It is easier to understand the concept of PCA if you visualise the

components without thinking about graphical representation. In the absence of graphical representation it is easier to imagine each of the PCs of the variance as being orthogonal (at right angles) to one another 'in space'.

For practical purposes, the first three or four PCs usually account for most of the variance. Getting up to the 11th and 12th PC, while they are still orthogonal to each other, the later PCs are accounting for much less of the true variance and become increasingly influenced by system noise, i.e. data that are not relevant to the measurement. In NIRS analysis, PLS regression includes both spectral and reference data and the principle is similar to PCR in that the factors are orthogonal to each other. As a general guideline, the fewer the PLS factors used in a model, the better. But it is seldom that a model uses only 3 or 4 and most PLS models use 6 to 8 factors. Caution must be observed if a PLS calibration uses only 1, or 15 factors (latent variables). It is impossible for a single factor (latent variable) to account for all of the combined spectral and reference data. But if the calibration uses 15 factors it is likely that more factors are needed, and that the software is considering that the later factors are essential. Such calibrations are unlikely to be of sustainable reliability.

Some software systems provide the option of displaying the weights across the spectrum. The weights are comparable to the loadings. A weights/loadings plot may appear as a spectrum, but it is not a spectrum. It shows the areas across the wavelength range where spectral variance has influenced computing of the model to a greater or lesser degree. The 'peaks' show the wavelength areas of the spectrum that has had the most influence on accounting for the variance. The spectral variance is influenced by the presence of absorbers such as O-H, N-H and C-H, which are recognised in NIRS work to be associated with water, protein, and oil, and 'bands' may be prominent in a display of the weights at wavelengths in the areas of 1900, 1400, 1154 and 960 nm for example. Both O-H and C-H absorbers occur in all of the major constituents of agricultural and food materials. The weights can best be used in conjunction with the information on the r^2-values for each PLS factor tabulated during the computation. Figure 10.6 shows the first two loadings for prediction of moisture content in whole-grain maize (corn). The 1st and 2nd PLS factors using log 1/R 'look' like the spectra of corn, but the strong influence in the wavelength areas associated with moisture is obvious. These two factors accounted for 82% of the total variance.

When moisture was predicted in the same sample sets using the 2nd derivative (2 4 4 1), the first PLS factor explained 93.6% of the variance, and the influence of water on the variance was more clear-cut (Figure 10.7). The 'bands' (wavelength

areas where the highest influence on variance was apparent) occurred at 1900, 1404, 1154 and 960 nm. Moisture content ranged from 13.8% to over 50%. An interesting point is that the 1st PLS factor for the prediction of moisture in wheat flour gave a similar display. Here the range in moisture content was only from 8.6 to 15.3%. The influence of water is apparent in many applications of NIRS to the analysis of grains (Williams, 2009).

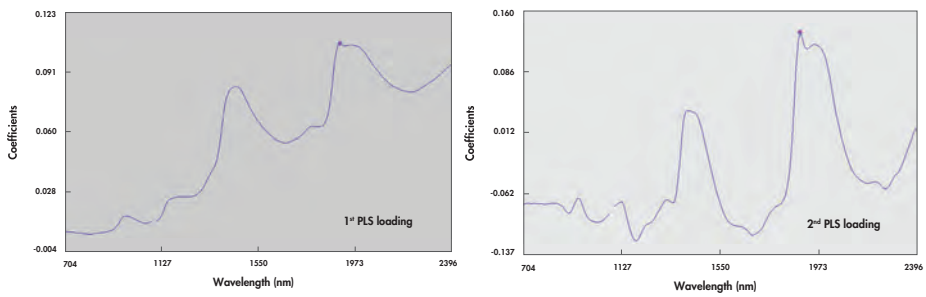


Figure 10.6 First and second PLS factors or loadings line plots for prediction of moisture in maize. Calibration has been developed with log 1/R spectra.

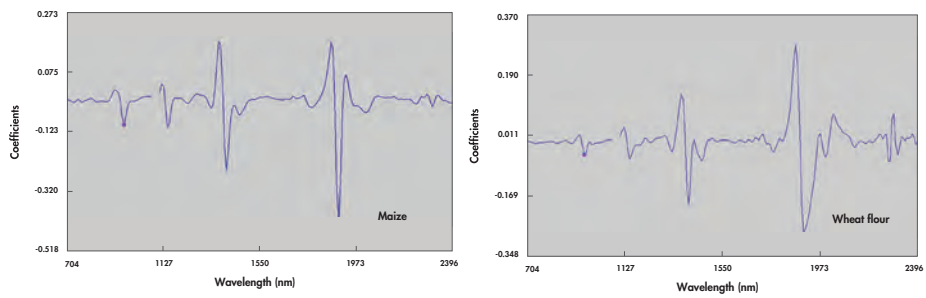


Figure 10.7 First PLS factors loadings line plots for prediction of moisture in maize and wheat flour. Calibration has been developed with 2nd derivative (2 4 4 1) spectra.

The interpretation of PLS weights becomes more complicated when NIRS is used for the prediction of functional parameters, such as wheat kernel texture. This is because unlike the NIRS prediction of constituents, such as protein, oil and moisture contents, there are no 'classical' absorbers assigned to functional parameters. The weights show the areas of variance that were most important in development of the model which can be used to speculate whether constituents, such as water or protein were active in model development. Figure 10.8 shows the first PLS factor computed during development of a PLS calibration model for prediction of kernel texture in ground wheat.

The log 1/R display looks like an 'upside-down spectrum' of wheat. The reference method for kernel texture is the PSI (particle size index) test. Wheat is ground to

whole-meal by a standardised method. An exact amount (10 g ± 1.0 mg) of whole-meal is sieved for 10 minutes by rotary action with percussion, using a 200-mesh sieve. The percentage of flour ('throughs') that pass through the sieve is recorded as the PSI. Hard wheat is more difficult to reduce to fine particles than soft wheat, and less flour is obtained after grinding than is obtained for soft wheat. Because of this, the PSI of soft wheat is much higher than that of hard wheat. The reason for the upside-down appearance is that the spectral data of hard wheat whole-meals show a higher degree of absorbance than do those of soft wheats (see Chapter 1, Figure 1.13). The absolute values of the reference data are much higher than the absorbance values, and are negatively related to the spectral data. The implication here is that the factor that exerts the most important influence on the PLS calibration is that of particle size. This is logical because of the known influence of kernel texture on the mean particle size of ground wheat.

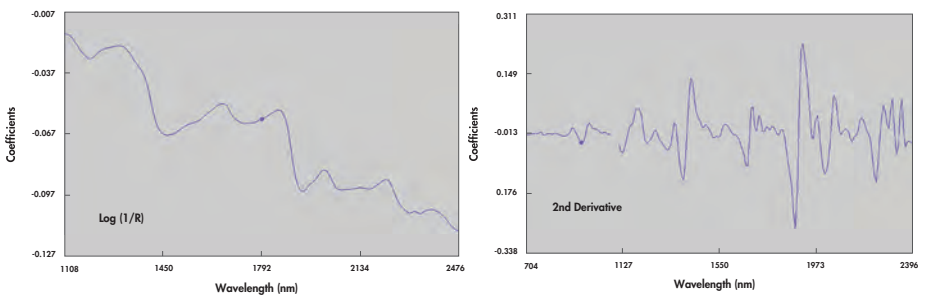


Figure 10.8 First PLS factor or loadings line plot for prediction of PSI in ground wheat. Calibrations have been developed with log 1/R and 2nd derivative (2 4 4 1) spectra.

The strong influence of variance in wavelength areas normally associated with water is again clear. The first PLS factor accounted for 91% of the total variance. The second PLS factor for prediction of ground wheat kernel texture (Figure 10.9) accounted for only 3.7% of the total variance, but the strong influence of variance in the main water band area around 1930 nm is obvious.

The first PLS factor usually accounts for the highest proportion of the total variance (i.e. the combined spectral and reference data) of the system. After the contribution of the first factor, the contribution of each individual PLS factor is accounting only for a proportion of the *residual* variance. An earlier paper (Williams, Cordeiro & Harnden, 1991) drew attention to the possible use of weights in interpretation of the process of model development. The paper showed that the strong influence of oil and protein on the prediction of dietary fibre components in oat bran was clearly revealed by the presence in the first few weights of strong peaks of variance in the 'classical' wavelength areas for these constituents.

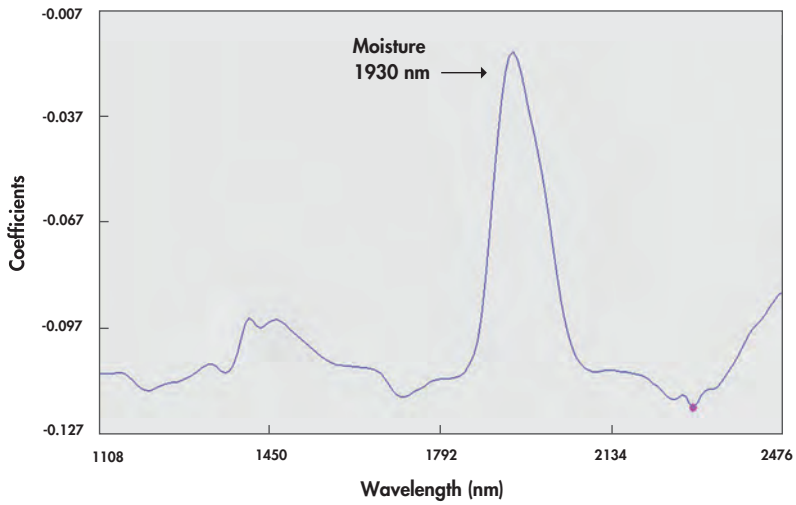


Figure 10.9 Second PLS factor or loadings line plot for prediction of PSI in ground wheat. Calibration have been developed with log 1/R spectra.

Table 10.1 Regression statistics for prediction of moisture in whole-grain barley

PLS factors	SEC	r^2-value	% Contribution*	F-value	SECV	1-VR	SEV	Bias
1	1.252	0.128	12.8	15.96	1.247	0.138	1.367	-0.085
2	1.209	0.186	5.8	8.18	1.211	0.186	1.298	-0.063
3	0.534	0.841	**65.5****	**413.32**	0.602	0.799	0.490	-0.111
4	0.498	0.861	2.0	16.15	0.509	0.856	0.434	-0.132
5	0.479	0.872	1.1	8.93	0.491	0.866	0.427	-0.135
6	0.383	0.918	4.6	55.47	0.411	0.906	0.313	-0.049
7	0.373	0.922	0.4	6.01	0.416	0.904	0.295	-0.056
8	0.369	0.924	0.2	2.99	0.422	0.901	0.277	-0.053
9	0.357	0.929	0.5	7.48	0.418	0.903	0.258	-0.041
10	0.338	0.936	0.7	11.73	0.410	0.907	0.251	-0.059
11	0.324	0.942	0.6	9.28	0.365	0.926	0.215	-0.052

*% Contribution to explaining overall variance; ** third PLS factor contributed most toward explaining the variance

The use of weights in interpretation of calibration models should be accompanied by the relative importance of each weight to the overall model. When developing a calibration model the r^2-values show how each successive factor is contributing to the explanation of the overall variance in the combined spectra and reference matrix (Table 10.1) as determined with WinISI software. Subtracting each r^2-value from the one succeeding gives the contribution of each factor to reducing the

residual variance. This is illustrated in column 4 (% Contribution) of Table 10.1, which is based on WinISI regression statistics. In the prediction of barley whole-grain moisture content, the third PLS factor contributed most toward explaining the variance in the system. Displaying the 3rd factor showed strong contributions to the variance at wavelengths 1410, 1340, 1156 and 964 nm all recognised 'water-bands' (Figure 10.10).

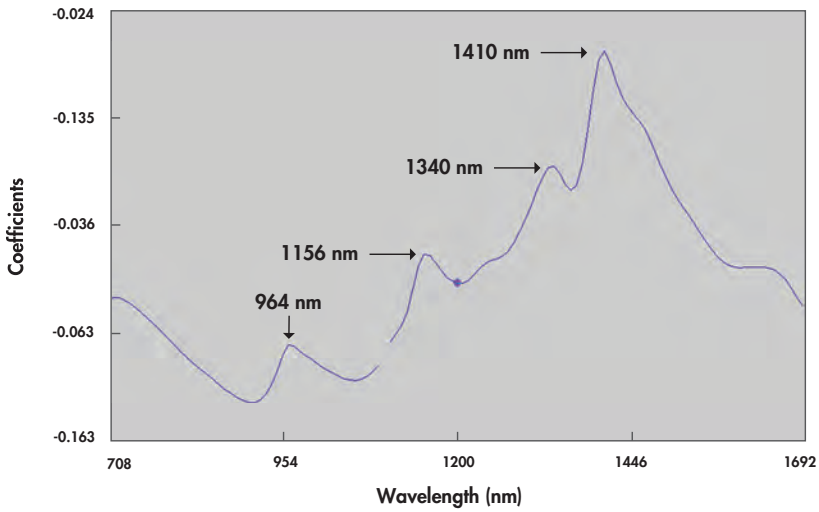


Figure 10.10 Third PLS factor or loadings line plot for prediction of whole-grain barley moisture content. Calibration have been developed with log 1/R spectra.

The wavelength areas where variance has apparently been used in computing the equation are of little practical value if the factor itself does not have a significant influence on the overall calibration model. This can be examined by studying the r^2- and F-values associated with each loading. Factors with the biggest F-values are the most important in the model development. The WinISI PLS option displays the F-values beside each PLS factor (Table 10.2).

Table 10.2 gives typical data for the calibration models for prediction of moisture content of maize and barley (both whole kernels) using PLS regression. In the prediction of maize moisture content, the first PLS factor accounted for nearly 94% of the variance in the system, and the F-values decreased very rapidly for the other factors. In the case of barley moisture content the third PLS factor had the most influence on the calibration model. The total amount of variance accounted for by the 11 factors computed, and the F-values were considerably lower and more variable. Table 10.3 gives some typical data for the degree to which individual PLS factors contribute to reduction in residual variance for some different parameters. The data for flour ash are unusual, in that the first three PLS

factors account for less than 20% of the overall variance. This observation was confirmed by repetition of the calibration process using different flour samples.

Table 10.2 F-values and relative contributions of PLS factors to calibration models for prediction of moisture content of whole-grain maize and barley

Moisture content of maize (%)				Moisture content of barley (%)			
PLS factor	**r^2-value**	**% Contribution***	**F-value**	**PLS factor**	**r^2-value**	**% Contribution**	**F-value**
1	0.936	93.6	4041	1	0.128	12.8	15.96
2	0.958	2.2	148	2	0.186	5.8	8.18
3	0.983	2.5	412	3	0.841	65.5	413.32
4	0.990	0.7	188	4	0.861	2.0	16.15
5	0.991	0.1	22	5	0.872	1.1	8.93
6	0.992	0.1	43	**6**	0.918	4.6	55.47
7	0.993	0.1	45	**7**	0.922	0.4	6.01
8	0.994	0.1	44	**8**	0.924	0.2	2.99
9	0.994	0.0	20	**9**	0.929	0.5	7.48
10	0.995	0.1	17	**10**	0.936	0.7	11.73
11	0.995	0.0	14	**11**	0.942	0.6	9.28
Prediction statistics				**Prediction statistics**			
r^2	0.993			**r^2**	0.919		
SEP (%)	0.62			**SEP (%)**	0.30		
Bias	-0.12			**Bias**	0.085		
Slope	1.010			**Slope**	1.011		
RPD	11.9			**RPD**	3.5		

*% Contribution to explaining overall variance (reduction in total and residual variance)

The PLS factors shown earlier are individual weights/loadings. After the first weight, each successive weight represents the residual variance, and even though a strong influence may be apparent in the displayed weight, the effect of that weight on the total variance in the system is tempered by the amount of the residual variance actually represented by the weight.

Table 10.3 Proportion of reduction in residual variance contributed by individual PLS factors

PLS factor	Proportion of reduction in residual variance contributed by each factor					
	Flour ash	Flour protein	Maize moisture	Flour water absorption	Wheat hardness A*	Wheat hardness B*
	I / C**	I / C	I / C	I / C	I / C	I / C
1	4.0 / 4.0	18.4 / 18.4	93.6 / 93.6	84.0 / 84.0	90.8 / 90.8	62.8 / 62.8
2	10.8 / 14.8	52.9 / 71.3	2.2 / 95.8	0.7 / 84.7	2.2 / 93.0	6.9 / 69.7
3	3.3 / 18.1	15.9 / 87.2	2.5 / 98.3	2.1 / 86.8	3.3 / 96.3	7.2 / 76.9
4	27.4 / 45.5	10.2 / 97.4	0.7 / 99.0	0.8 / 87.6	0.5 / 96.8	2.7 / 79.6
5	3.2 / 48.7	1.9 / 99.3	0.1 / 99.1	1.7 / 89.3	0.4 / 97.2	1.8 / 81.4
6	24.7 / 73.4	0.1 / 99.4	0.1 / 99.2	3.1 / 92.4	0.5 / 97.7	2.3 / 83.7
15	1.3/94.2	0.1 / 99.8	0.1 / 99.5	0.5 / 98.8	0.2 / 99.0	1.7 / 95.0

*A = ground kernels; B = whole kernels; **Proportion of reduction in total variance contributed by individual factors expressed as I / C = percent of individual / cumulative (total) variance explained

10.4.2 PLS regression coefficients (b)

Regression vectors are the regression coefficient (b) developed at each wavelength point. Using PLS regression, they can be displayed in the same way as the weights. The regression coefficients (b) are the result of the combined influence of all of the PLS factors. Because of interactions among other constituents this can result on a confusing array of apparent influences, as shown in Figure 10.11. Contributions to explanation of the variance at individual wavelength areas can be positive or negative, and the assigning of an absorber to an individual peak or valley should be made with caution.

The variables that affect the spectral characteristics of a material include its physical, as well as its chemical make-up. The display of regression or b-coefficients shown in Figure 10.11 were for the prediction of moisture in whole maize kernels which vary widely in size, packing density and the consequent degree of scatter. If NIRS is applied to a material, such as wheat flour, which is much more uniform in particle characteristics, the b-coefficients for moisture prediction show a clear effect in the wavelength areas associated with O-H absorbers (Figure 10.12). The arrowed regression coefficients (b) in both Figures 10.11 and 10.12 indicate the strong water band at 1930 nm (Workman and Weyer, 2012).

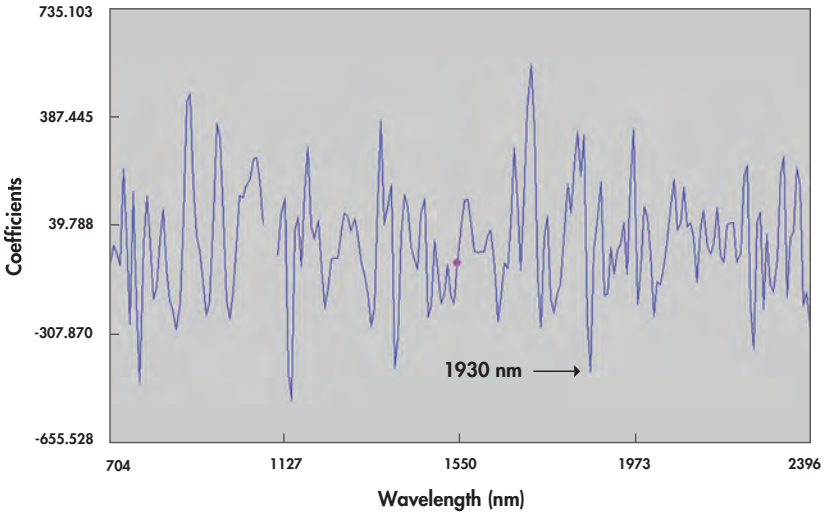


Figure 10.11 Regression coefficients (b) for prediction of moisture content of whole maize. Calibration has been developed with 2nd derivative (2 4 4 1) spectra. The 'valley' at 1930 nm can be assigned to O-H (water) which absorb at that wavelength.

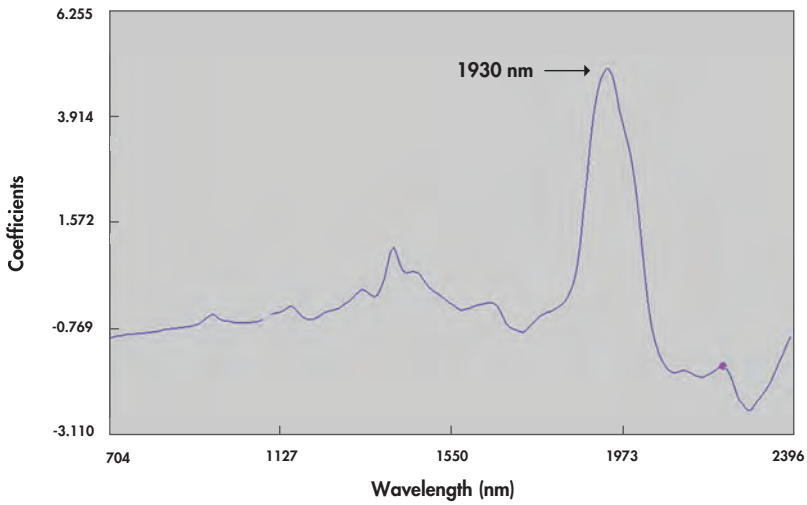


Figure 10.12 Regression coefficients (b) for prediction of wheat flour moisture. Calibration has been developed with log 1/R spectra.

For each application of NIRS to different materials and prediction of different parameters the PLS factors and b-coefficients give some information on the interactions that have taken place during the computing of the calibration model. As a general rule the PLS factor or loading line plots, in combination with the r^2- and F-values, give a better insight to the development of the calibration model than do the b-coefficients.

10.5 Assignment of wavelengths to constituents

Classical infrared spectroscopists have mapped the position of the fundamental absorbers for the C-H, O-H, N-H, C-N, C-O and other functional groups in the mid-IR region. All absorbances in the NIR wavelength range derive from these fundamental absorptions. Because of the multitude of absorbers that occur in the NIR wavelength range, extensive overlap of absorbers occurs which complicates making definite wavelength assignments. For example, in the immediate area of 1940 nm there are no less than ten possible absorbers, including combination bands, and overtones involving C-H, O-H and C-O. In biological (including agricultural and food) materials, including methyl (CH_3) methylene ($-CH_2$) and aromatic C-H groupings there are more absorbers associated with C-H than with other molecular arrangements. This is because all of the major constituents, starch, cellulose, protein and oil/fat are all rich in C-H molecular groupings, such as -CH-X, $-CH_2$ and $-CH_3$. In working with materials and derived materials in the agricultural, food, forage and feed domains it is more practical to assign an absorber to a constituent, such as protein or water, rather than to a specific functional group. This mass of absorbers is an asset to quantitative analysis. When absorbers interfere with each other at some wavelengths, the computer is able to select appropriate absorbers from other wavelengths that are more highly correlated with the reference data.

Figure 10.13 shows the approximate positions of the main constituents of agricultural materials on the spectrum from 570–2500 nm. The actual spectrum is of ground oats, but the positions are applicable to many agricultural commodities.

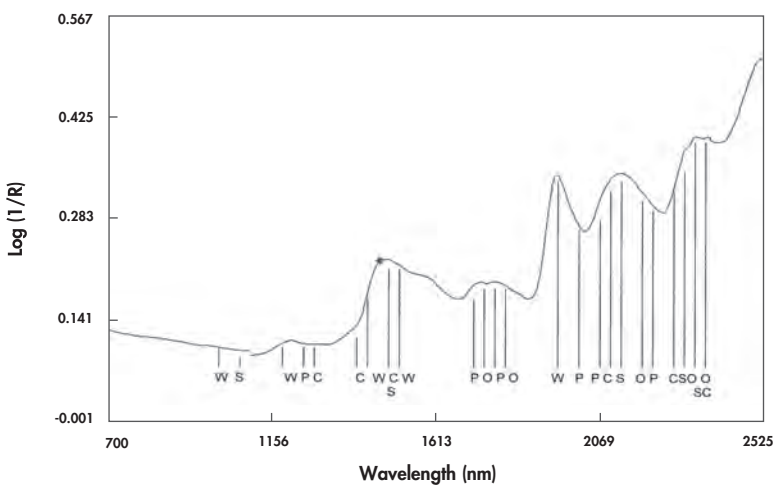


Figure 10.13 Spectrum of ground oats showing areas where major constituents absorb (C = cellulose or fibre; O = oil; P = protein; S = starch; W = water).

The main wavelengths where the most important components (in terms of composition) of agricultural materials have been found to absorb, are shown in Chapter 2. Table 2.2 is included to help with identifying what is influencing calibration development in this way. These have been determined by preparing the constituents in as pure a form as possible from wheat and canola seed (oil only). The spectra were recorded on a computerised spectrophotometer, scanning from 400–2500 nm. The mathematical treatment of 2 4 4 1 was applied to the spectral data to 'sharpen' the bands. It is particularly interesting to identify the wavelengths associated with functional parameters, such as wheat kernel texture and wheat flour 'strength'. Full details of absorbers in the NIR region are given in Workman and Weyer (2008) (see *Suggestion for further reading*).

Reference

Williams, P.C., Cordeiro, H.M. & Harnden, M.F.T. 1991. Analysis of oat bran products by near-infrared reflectance spectroscopy. *Cereal Foods World*, 36, 571–574.

Williams, P.C. 2009. Influence of water on prediction of composition and quality factors: The aquaphotomics of low moisture agricultural materials. *Journal of Near Infrared Spectroscopy*, 17, 315–328.

Suggestion for further reading

Workman Jr., J. & Weyer, L. 2012. *Practical Guide and Spectral Atlas for Interpretive Near-infrared Spectroscopy.* 2nd Edition. Boca Raton, Louisiana, USA: CRC Press, Taylor and Francis Group.

11

Applications of NIRS

The message of successful applications of NIRS in agricultural fields has attracted the attention of the pharmaceutical, petrochemical and other major industries. The early diagnosis of diseases in animals and humans is receiving increasing attention. Other important applications include investigations into environmental change. The applications of NIR technology are far too extensive to cover in this chapter. The excellent handbooks compiled by Don Burns and Emil Ciurczak (2007) and by Craig Roberts and co-authors (2004) (see *Suggestions for further reading*) cover many applications. These include applications in the areas of agriculture, food, industrial processing, medical, pharmaceuticals, petrochemicals, textiles and others. The principles covered in this chapter are, however, applicable to any type of NIRS application.

11.1 What can be measured with NIRS?

The most frequent applications are for determination of composition, particularly moisture, protein, oil/fat, fibre and starch contents. Calibrations are also available for factors such as individual amino acids and fatty acids. Texture (the physical structure of materials) and particle size affect light scatter, and because of this it

is possible to measure factors such as degree of hardness/softness in grains, an important factor in grain processing, soil particle size (composition relative to the contents of sand and clay) and the viscosity of slurries, such as liquid manures. Minerals, such as calcium and phosphorus, can be measured because of the way in which they occur in plant materials. Some other elements can be measured in materials such as liquid manures because of the form in which they exist, and by their association with other constituents.

Advances in data analysis software have made it possible to obtain estimates of some quality characteristics in grains, such as wheat gluten strength, barley malt hot water extract, digestibility in forages and digestible energy in feed grains that are sufficiently reliable for use in plant breeding. Elements such as nitrogen, phosphorus, carbon and other elements can be measured with accuracy in manures and soil. Qualitative analysis, used in the form of classification analysis, is employed in the pharmaceutical industry to identify the substances used in prescription and other types of medicinal drugs on the basis of the relatively simple chemicals used in their manufacture, as well as to verify that the composition is within specifications.

What *cannot* be measured reliably with NIRS? In general, constituents such as trace elements, vitamins, pesticide and herbicide residues cannot be reliably predicted by NIRS. Although there are millions of molecules present, they are present in very small amounts relative to other constituents with the same functional groups. Enzymes are proteins and most of them cannot be measured reliably because of their general similarity to other proteins that are also present in far greater quantities. Some enzymes, such as alpha-amylase and some mycotoxins, such as DON, a mycotoxin of *Fusarium* species that causes 'scab' in wheat can be estimated indirectly because of their influence on grain texture. Heavy metals cannot be predicted reliably using NIRS unless they are strongly associated with a constituent that can be predicted accurately, such as cellulose (Malley & Williams, 1997).

Constituents that are present in very low amounts are all made up of the same basic molecular groupings (C-H, O-H, N-H) as the major constituents. There are many more of these absorbers present in the major constituents of a material and these will dominate the spectra. As the instruments become even more precise than they are now it may be possible to obtain estimates of constituents that are present in low concentrations, but because of the way the system works, NIRS is not likely to take the place of, for example, atomic absorption spectrophotometry for the determination of heavy metals, in terms of accuracy. Similarly, NIRS

estimates may be possible of organoleptic qualities, such as flavour, but again, the differences in flavours are affected by constituents that are present in extremely small concentrations, and often more than one of these interact with each other to provide the subtle changes in taste that can be detected reliably by trained taste panels.

It is important not to expect too much from an NIR instrument. There are many instances where a company, a university, or some other organisation buys an expensive instrument and is disappointed and disillusioned when their instrument fails to give them the results that they expected. Modern NIR instruments are capable of a very high degree of spectral precision. The burden of the effectiveness of their application becomes the responsibility of the analyst. In many cases the main reason for mediocre results can be traced to mediocre reference analysis. In some cases, the application fails because there are simply no absorbers that relate to the parameter to which the technique is applied. But in many other cases indifferent results occur because the operators do not understand the technology, do not know how to use the instrument properly, and as a result are unable to exploit its benefits. These situations reflect unfairly on the instruments and give NIRS a bad name. Hence one of the main reasons for this book.

A very important aspect of application to agricultural and food materials is the method of sample presentation to the instrument because of the diversity of physical form of the materials, particularly fresh materials, such as meat, fish, fruit and vegetables.

This chapter is limited to applications in the agriculture, food and feed industries. These applications are dealt with alphabetically and not in order of importance, although there are probably more instruments applied in grain testing than in any other field.

11.2 Feed industry

Analytical chemistry is becoming increasingly important in the animal feed industry. The constituents most frequently sought are moisture, protein, fibre, fat and ash contents. NIRS can save large analytical costs and because of the short analytical time, it can also extend the frequency and flexibility of testing. Feed manufacture is a continuous process, and the on-the- spot results provided by NIRS is a valuable asset to quality control in a feed mill, as well as the prevention of costly down-time. The NIRS predicted results should be related to the tolerance levels and specifications demanded by the industry, and also to the relative

precision (reproducibility) of the laboratory and NIRS methods of analysis. In many cases the precision of the NIRS instrument is superior or at least equal to that of the laboratory, and the time saved by using NIRS can be a valuable incentive to use it in quality control.

Most industrial applications deal with volume and small benefits and small errors can translate into big money. Profit margins are often small, and differences in the price per tonne (metric tonne) of grain of only one or two dollars per tonne can translate into hundreds of thousands of dollars over an annual production cycle. Protein supplements such as soybean meal are expensive, and NIRS analysis in the feed mill can enable feed manufacturers to derive the most efficient feed formulations.

More attention is being paid to the advantages of formulating feeds on the basis of composition and metabolisable energy. Efficient feed formulation results in maximum productivity in terms of weight gain, egg and milk production and other areas. Metabolisable energy testing requires feeding animals on controlled diets for several days. The costs are prohibitive and complicated by the fact that large numbers of samples representing the range in energy are required, so that a) assembling a sample set is very time-consuming; b) finding a facility to carry out the testing is difficult; and c) nutritionists are not in complete agreement as to which version of metabolisable energy is the right one to use. AME differs from TME, and the most appropriate form depends on the animal.

Nevertheless, metabolisable energy is fairly highly heritable in grains and a calibration model has been developed for prediction of TME in barley. This has great potential for barley breeding programmes and also in feed formulation. For the development of this type of calibration and the length of time of the reference methods, sample assembly can take more than a year. Gross energy can be determined using a calorimeter and is easier and cheaper to determine, but of less value in feed formulation. Table 11.1 summarises results obtained for NIRS prediction of some fibre components in feed barley. The time and costs per test for reference and NIRS testing are included. Total time for developing the TME calibration is not included in Table 11.1, but it took over three years to accumulate the >100 samples used to develop the calibration, and only one day to scan all of the samples and develop the calibration model! The difference in analytical costs is striking, but realistic.

Table 11.1 Near-infrared analysis of feed barley

Analysis	r^{2a}	SEP[b]	Bias[c]	b[d]	RPD[e]	Time per test		Cost per test ($US)	
						Reference	NIRS[f]	Reference	NIRS[f]
NDF (%)	0.978	0.434	0.108	1.002	6.7	12 h	2 min	99	12
ADF (%)	0.973	0.326	0.097	1.011	6.0	12 h	2 min	99	12
CF (%)	0.973	0.255	0.031	1.035	5.9	12 h	2 min	99	12
TME (units)	0.851	0.259	0.097	0.972	2.6	10 days	2 min	650	12
Total	-	-	-	-	-	10 days	2 min	947	12
Cost for calibration model development (100 samples)								94,700	1,200

For definitions see the following sections in Chapter 4: a = 4.1.6; b=4.1.10; c = 4.1.4; d = 4.1.7; e = 4.1.13; f = all four tests can be carried out simultaneously

11.3 Food industry

Applications in the food industries include baking, breakfast cereals, cake mixes, dairy products, fish, meat, noodles, oils and sauces, pasta and others. The statements on packaging are legal obligations and apply in most countries. Processing plants need to know that the correct ingredients are present in the correct proportions, and that the final products meet consumer and legal demands. The manufacturer must be prepared to answer legal questions should they arise. Modern NIR instruments are capable of degrees of accuracy and precision equal to those of 'wet chemistry', and factory calibrations are offered for application to a wide range of foods. Constituents that are most frequently sought include egg content, energy, fat, fibre (crude and total dietary fibre), protein, starch, sugar and water contents, all of which can be reliably predicted by NIRS. The reference made of the advantages of using NIRS in quality control (see section 11.2) is equally applicable in the food, and indeed in most industries that call for analytical work.

11.4 Flour milling

The flour milling industry has employed NIRS for almost as long as it has been used in grain-handling. Flour is usually marketed to bakers, noodle-makers and pasta-makers on the basis of specifications. These involve protein and ash contents, and often water absorption and Farinograph or Alveograph parameters. Some mills monitor starch damage, which can also be predicted with NIRS. For the actual milling process, millers need to know the moisture, ash and protein contents of the final products, and of mill streams as the millers have found to be useful in

monitoring the continuous performance of the mill. Where mills purchase wheat directly from farms, NIRS is used to test for protein content at the time of delivery. Grists can be blended to assure that the end-product flours meet specifications for protein and ash contents, so the efficiency of blending is also important.

The Buhler Corporation offers on-line NIR instruments that continuously monitor possible fluctuations in moisture, protein and ash contents of the flour. Collaborative studies carried out on Approved AACCI Method 08-01.01 for ash determination (AACC, 2000a) and Method 08-21.01 for ash determination using NIRS (AACC, 2000b) showed that for most collaborators the precision (as measured by the standard deviation between blind duplicates) was slightly better for the NIRS method than for the incineration method. The respective mean SD (standard deviation) of duplicates for NIRS and the incineration method were 0.0048 and 0.0083, and the respective CVs were 0.87 and 1.51% (see Chapter 4, section 4.1.3 for definition of CV). The NIRS ash test takes about one minute, and protein and moisture contents can be determined at the same time, whereas the ash reference test takes at least 2.5 h, and is less precise than the NIRS method. The NIRS system can be set up for continuous monitoring of these three constituents. A network of two or more instruments can be set up to monitor flour production at critical points, all controlled from a single desktop or laptop computer. Flour quality parameters that have been predicted by NIRS are summarised in Table 11.2.

Table 11.2 Application of NIRS to prediction of quality factors in flour

Test	r^2	SEP	Mean	SD	RPD	Potential
Moisture (%)	0.99	0.187	12.78	1.8	9.6	Excellent
Protein (%)	0.995	0.105	12.7	1.5	14.1	Excellent
Flour ash (%)	0.88	0.016	0.55	0.05	2.9	Good
Flour starch damage (Megazyme units)*	0.93	0.38	5.82	1.47	3.9	Very good
Farinograph water absorption (%)	0.92	1.32	63.4	4.5	3.4	Very good
Dough development time (min)	0.58	1.40	5.2	2.1	1.5	Poor
Dough stability time (min)	0.68	2.78	7.8	5.0	1.8	Poor
Mixing Tolerance Index	0.92	14.1	52.4	51.0	3.6	Very good
Alveograph W	0.98	23.4	237	161	6.9	Excellent

The determination of flour ash by NIRS is an Approved Method of the AACCI (AACC, 2000b). Flour ash is conventionally determined by incinerating the flour and weighing the resulting ash (reference method). The ratio of ash (actually mineral content) in wheat bran to that of the endosperm (the eventual flour) is about 29:1, so a small admixture of bran will increase the ash content of the flour significantly. The ratio of fibre in wheat bran, relative to the endosperm is about 25:1. Cellulose is a major constituent of bran, and the correlation coefficient (r) between flour ash and flour cellulose is over 0.9. There is no ash in flour until it is incinerated, so the NIRS ash method probably works by predicting 'ash' on the basis of changes in cellulose content. The cellulose would arise from small inclusions of bran during milling.

11.5 Forage analysis

It was in the application of NIR technology to forage analysis that the first attempts at network development were made. This was in a collaborative effort in the mid-1970s between USDA scientists, mainly at Athens, GA, USA, and Pennsylvania State University, home of the late Dr John Shenk, one of the pioneers in application of NIRS. The biggest difficulty with forage analysis, including forage at all stages of growth and straws, lies in sampling, sample preparation and presentation to the instrument. Probably the best way to sample a large field of mature forage or straw is to sample bales from different spots in the field. Core samples can be taken from the targeted bales and carefully sub-sampled before sample preparation. If the bales are numbered and their locations in the field recorded, the field can be mapped according to constituents such as nitrogen and phosphorus.

Fresh, immature forages are usually not baled and samples must be taken of the actual forage on the spot. Analysis of fresh or mature forages/straws in the long state (before size reduction) is not practicable. Dried forages are quite light, and even the Foss 'natural products' rectangular sample cell would only be able to accommodate samples from a very small area of a field or plot. Because the material has to be chopped, and/or dried before sample preparation, determination of moisture of fresh forages is essential before and after sample preparation for purposes of model development. The true value of the forage as feed depends on the state at which it is ingested. Forage that has 24% protein on a dry matter (DM) basis, but only 18% DM at the time of ingestion by the animal, has only 4.3% actual protein. This, together with the estimated amount of forage ingested, is the value that has to be used to compute any supplements that are to be fed to maximise production, and also the value to be used in computing a calibration model for the fresh forage. The amount of forage eaten by the animal has to be taken into

consideration. Voluntary intake (actual consumption of feed by the animal) would be a welcome addition to the parameters measurable by NIRS.

Parameters in forages for which NIRS calibration models have been developed include dry matter, protein, fibre components, such as NDF, ADF, crude fibre, *in vitro* and *in vivo* digestibility, lignin, dry matter intake and some ash components. Recent advances in instrumentation have resulted in an NIRS instrument being attached to a forage harvester. This technique is called on-site NIRS. It allows the forage to be tested for moisture content continually during harvest. This is an important advantage to farmers, who get paid on a dry matter basis for fresh material delivered for the production of bio-gas. Figure 11.1 shows the instrument in place on the spout of a forage harvester, and in action in the field.

Figure 11.1 An NIRS instrument on the spout of a forage harvester (left) and in action in the field.

11.6 Fruit and vegetable applications

Applications of NIRS in the fruit and vegetable industries include the determination of °Brix (a measure of sweetness), moisture content, degree of ripening, firmness and organic acids. Interior defects can be detected in vegetables such as potatoes, while they have also been successfully predicted in apples and some stone-fruits. The main difficulties with application of NIRS to fruit and vegetables is sample presentation to the instrument, and presentation of the instrument to the sample.

11.7 Grain handling

The aspects of grain handling to which NIRS can be applied include composition analysis for constituents such as moisture, protein, starch and oil contents. Other applications include classification into grains of different types on the basis of their functionality, and grading into different categories within a grain type, on the basis of damage that has been caused to the grain by weather during maturation

and post-harvest, storage and transportation. This is a qualitative rather than a quantitative approach.

11.7.1 Composition

Of all of the applications of NIRS in grain-handling, the most widely-used is composition analysis. During the first half of the 1970s analysis of wheat for protein content by NIRS brought about a minor revolution in domestic and foreign wheat marketing from the point of delivery in the country to the elevator (silo), flour mill and shipping port. This was quickly extended to barley, although world-wide, wheat is by far the most important grain commodity traded. About 30 years ago ANN calibrations were introduced for moisture and protein contents that are applicable to both wheat and barley.

11.7.2 Classification

Classification of grains into different categories on the basis of their end-use is practiced in some form by all of the major exporting countries. In Canada, the USA and Australia, examples are classes of wheat such as Dark Northern Spring and Soft Red Winter (USA), CWRS (Canada Western Red Spring), CPSR/W (Canada Prairie Spring Red and White), CWES (Canada Western Extra Strong), Australian Standard White (ASW) and Australia Prime Hard (APH).

In Canada and the United States, the classes of wheat are bred to carry kernel characteristics that enable grain inspectors and elevator managers to distinguish them visually from one another. Prices vary, depending on the class and grade within a class. In Australia, at least in New South Wales, protein content and kernel hardness (based on wheat variety) make up the principal system for separating wheat into classes, and NIRS is used for identifying wheat by class and within class (pricing level). To date, NIRS is not used in North America for classifying wheat, but because of significant premiums paid for protein content, up to US$1.00 per 0.1% protein above 12% in CWRS wheat, NIRS is an important factor in assessing price to the farmer, and to elevator companies. Classification will be discussed later in this chapter.

11.7.3 Grading

Weather conditions during the growing and postharvest periods, infestation by fungi, harvesting methods and transportation can all affect the appearance and to varying degrees the functional properties of grains. These factors have all been studied, as well as their impact on the grains incorporated into systems for grading

and pricing the grains. Weather conditions, such as frost during the maturation phase and post-harvest sprouting induce changes in grain texture, and therefore the spectra. Because of these changes NIRS should be applicable to detection of these grade factors.

Frost tends to cause kernel texture to become denser, and the kernels become harder than normal, which affects milling quality. Factors that induce bleach, mildew and sprouting, such as postharvest wetting and drying in-swathe, reduce the density of the wheat. A lot of the applied research on wheat has been carried out by industrial organisations whose main objective is to improve their operation by extending NIRS (and other techniques, such as the Perten SKCS - Single Kernel Charactersation System), to aspects of grain-handling. As a result, a great deal of the work has remained unpublished. Some of the work has been induced by 'knee-jerk' reactions to seasonal conditions that show up as soon as the new crop reaches the market place. Work carried out in Canada on the prediction of wheat grading factors is summarised in Table 11.3.

Grading by NIRS would probably lead to an overall improvement in the precision of grading. This could be assisted by use of hyperspectral imaging. At present, introduction of new classes of wheat is exerting pressure on the visual grading system, mainly because the differences in kernel characteristics between some classes, such as the Canadian CWES and CPSR classes, are subtle. Most of these classes differ in functional and end-use potential.

Table 11.3 Application of NIRS to visual wheat grading factors

Grading factor	r^2	SEP	Mean	SD	RPD	Potential for use
Test weight (wheat)	0.79	0.77	82.29	2.14	2.8	Good
Falling Number	0.62	42.5	334	54	1.3	Poor
FDK (%)	0.69	0.81	1.80	1.95	2.4	Poor
Frost (%)	0.82	4.62	10.1	19.5	4.2	Good
HVK (%)*	0.77	5.4	70.5	15.1	2.8	Good
Plump kernel (%)**	0.83	11.5	59.5	28.1	2.5	Fair
Chlorophyll (ppm)	0.96	2.06	30.9	10.8	5.2	Excellent

FDK = *Fusarium*-damaged kernels; HVK = Hard vitreous kernels; *only in durum wheat; **only in barley

11.8 Hyperspectral imaging

Near-infrared hyperspectral imaging (NIR-HSI) is a relatively recent introduction to food and agriculture applications (Manley, 2014). The technique involves taking images of materials and objects, dividing the images into a series of sub-images called pixels, and developing an NIR spectrum for each pixel. The number and complexity of spectra data recorded depend on the size of the array photographed, and on the detector used. Even with silicon detectors, over the small wavelength range of 700–1100 nm the database becomes very big, and considerably bigger with an InGaAs detector, with a wavelength range of 900–2500 nm.

The equipment is expensive. It includes a digital camera, equipped with a built-in spectrometer with the appropriate detector, and associated software. The system is operated with a computer. The database is very big, which also affects computing time. The NIR-HSI technology is at present most suited to qualitative classification (discriminant) analysis applications. It appears to be the ideal system for electronic grading of materials such as grains or meat. The problem of sample presentation to provide a representative image of a large population of grains appears to have been overcome, but research is continuing on the question of translating the enormous amount (terabytes) of data into industrially acceptable procedures for grading or classifying grains, meat, fish, fruits and other commodities and products.

11.9 Manure analysis

Manure is as much an agricultural product as grain or hay. Manure management is receiving a great deal of attention world-wide, although up to the present very little analysis is being carried out on the industrial basis. The practice of spreading animal waste on the land can lead to serious pollution of surface and ground water systems and excessively high levels of soil nitrogen and disproportionate amounts of phosphorus below the growing zone. The disposal and/or mismanagement of manure has the potential to contaminate water supplies to cities, small towns, and particularly villages in areas where hog production is high. This can also cause human disease, cause intense eutrophication (increase in chemical nutrients, particularly nitrogen and phosphorus) of streams, and fish kills; and change the species composition of algae in rivers, lakes and oceans adjacent to major hog producing areas.

From the farmers' viewpoint there is a need to know what nutrients are being applied to their land, both from the aspects of the degree to which they may need to supplement the manure with commercial fertilizer to meet crop requirements,

and the degree to which they are meeting restrictions on nitrogen and phosphorus application. Hog manure has commanded the most attention, due to the way it is often stored, difficulties in handling and its high phosphorus content. In North America, it is often stored in tanks, or earthen ponds or lagoons (Figure 11.2), which can hold 4 to more than 20 million litres of liquid manure. A significant amount of the material in the manure is undigested food. Surface layers of the stored manure may contain as low as 0.5% dry matter and lower layers up to over 11% dry matter. The solid material tends to settle, so that the manure is very variable in composition, and sampling has to be carried out carefully in order to represent the whole lagoon. Ideally, the manure is analysed continuously as the manure store is pumped into tanker trunks or applied directly to fields. The main constituents of concern are nitrogen (N) and especially phosphorus (P). Both of these, as well as potassium, ammonia, and carbon, can be determined in liquid manure by NIRS.

Figure 11.2 A hog manure lagoon before (left) and after pump-out.

NIR instruments can provide continuous monitoring during the pump-out and provide a good estimate of the total amounts of N (nitrogen) and P (phosphorus) that has been spread on the land. Manure management is a potentially multimillion-dollar industry, which can also improve the efficiency of fertilizer application and save considerable expense to grain farmers in the area. The objective is to apply N and P in a balanced ratio that will be utilised as fertilizer by the crop. On the average, hog manure has an overabundance of P relative to the amount of N. Analysis by NIRS during lagoon pump-out can be used to assist in applying the correct amount of P to the field and indicating the amounts of additional N fertilizer that need to be applied to achieve the optimal balance between nitrogen and phosphorus.

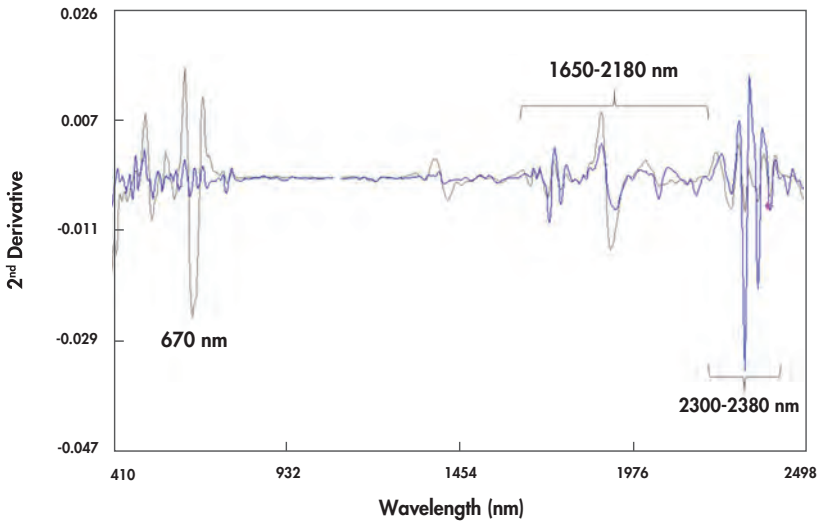


Figure 11.3 Dairy cow (light) and hog (dark) manure 2nd derivative (2 4 4 1) spectra.

Figure 11.3 shows 2nd derivative (2 4 4 1) spectra of dairy cow and hog manure. Several features are apparent. First, notice the strong band in the 670 nm area in the cow manure due to chlorophyll that gives manure from pasture-fed cows the familiar greenish colour. The band is absent from the hog manure. Next, note the very prominent bands at 1726, 1762, 2052, 2168, 2308 and 2350 nm in the hog manure spectrum, most of which are far less prominent in the dairy cow manure. These are caused by residual protein and fat in the hog manure. Hog nutritionists can use data of this type to assist in formulation of efficient diets. The data indicate that the hogs were not digesting their feed very efficiently. For example, phytates

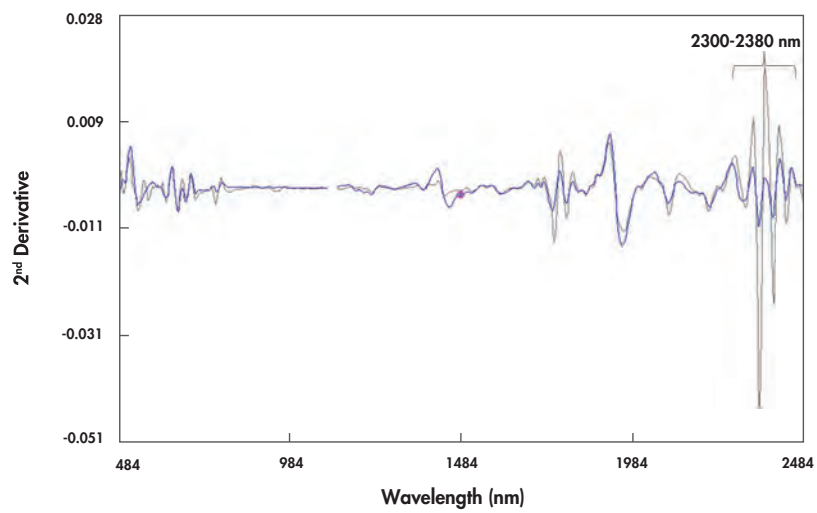


Figure 11.4 Hog (light) and broiler (dark) manure 2nd derivative (2 4 4 1) spectra.

in the feed grain are poorly digested by hogs since the animals lack the enzyme phytase. This leads to large amounts of residual phosphorus in the manure. Some hog-producers use phytase in the diet. This improves the digestibility of phytate, and reduces the phosphorus content of the manure. Figure 11.4 compares the 2nd derivative (2 4 4 1) spectra of hog and broiler (chicken) manures. While the bands in the protein and oil band areas (2300–2380 nm) are prominent in the broiler

11.10 Meat and fish analysis

Analysis of fish for fat and protein contents dates back to 1987 (Mathias, Williams & Sobering, 1987). Since then NIRS has been applied to the analysis of fish to determine fat content and the presence of parasites in dead, frozen and live fish. As well, changes in muscle development have also been studied, using NIRS. The applications of NIRS to meat analysis concern mainly determination of fat content, and more recently water holding capacity or 'drip loss'. Chicken carcasses are examined on-line by NIRS for contamination with faeces, using fibre optics.

11.11 Mixing and blending efficiency

The efficiency of mixing and blending in flour and feed mills and other processing plants is sometimes tested by sending samples for analysis of a minor constituent before and after the mixing-blending operation. This method is subject to the errors in the analytical tests used, as well as to delays in obtaining the analyses. The efficiency of mixing/blending can be determined simply and quickly by using NIRS spectral data with no chemical analysis. The first step is to determine what wavelengths to monitor. It is a good idea to monitor wavelength points that are known to be associated with known constituents, such as water (982, 1162 and 1412 nm), oil (1210 and 2306 nm) and protein (1692 and 2054 nm).

Samples (at least 10) are taken and scanned after about five seconds of blending and again after blending is believed to be complete. If the blending is efficient, the SD of the spectral peak heights after five seconds of blending will be several times greater than the SD of the peak heights at the same wavelength points after blending is complete. Table 11.4 shows the influence of blending on the spectral data of wheat flour at two wavelength points. For brevity, only five replicates are shown. The variance in the log 1/R of the unblended series at both 1410 and 2306 nm was about seven times higher than that of the blended series. This showed that monitoring the variance in log 1/R provides an excellent estimate of the efficiency of blending.

Animal feeds usually contain several ingredients, most of which differ considerably from each other in physical characteristics. Applications in the feed industry include analysis of raw materials, feed-mix formulation, analysis of the final mixes and testing mixing efficiency. Practically all domestic farm animal feeds are based on grain. The grain that forms the basic ingredient varies depending on price. Other ingredients include protein and mineral supplements, fibre sources, such as alfalfa pellets, meat and bone meal and fish-meal depending on local availability. Because of major differences in size and texture in their original form, these are all ground in a hammer mill before mixing, but still vary widely in their spectral signatures.

Table 11.4 Changes in precision of Log 1/R values as a result of blending flour

Replicate	Log 1/R at 1410 nm	Log 1/R at 2306 nm
1a*	0.3376	0.5318
1b	0.3364	0.5316
2a	0.3344	0.5277
2b	0.3369	0.5325
3a	0.3373	0.5329
3b	0.3369	0.5322
4a	0.3425	0.5423
4b	0.3376	0.5336
5a	0.3362	0.5303
5b	0.3371	0.5328
SD_a	0.00301	0.00555
Mean log $1/R_a$	0.3376	0.5330
CV_a	0.892	1.042
SD_b	0.000432	0.00074
Mean log $1/R_b$	0.3370	0.5325
CV_b	0.128	0.139
Ratio CV_a/CV_b	6.97	7.50

*Series 'a' samples were unblended; series 'b' samples were thoroughly blended

The same system (as the one used in the above example) of testing the efficiency of blending flour mill streams can be applied to monitoring mixing efficiency in a feed mill. Here the mixing is complicated by the fact that the ingredients used in a feed mix are often quite different in physical makeup and spectral characteristics

and efficient mixing is essential. The respective SDs and corresponding CVs are determined of samples that are withdrawn and scanned early in the mixing and after mixing. Figure 11.5 shows the dramatic reduction in CV in a dairy cattle feed mix as a result of efficient mixing. The upper line in Figure 11.5 shows the CV in spectral data across the spectrum before mixing and the lower line shows the CV after mixing. The reduction in the variance in spectral data was dramatic, and gives the operator the ability to determine the efficiency of mixing/blending within a few minutes, and with no need for laboratory or chemical analysis.

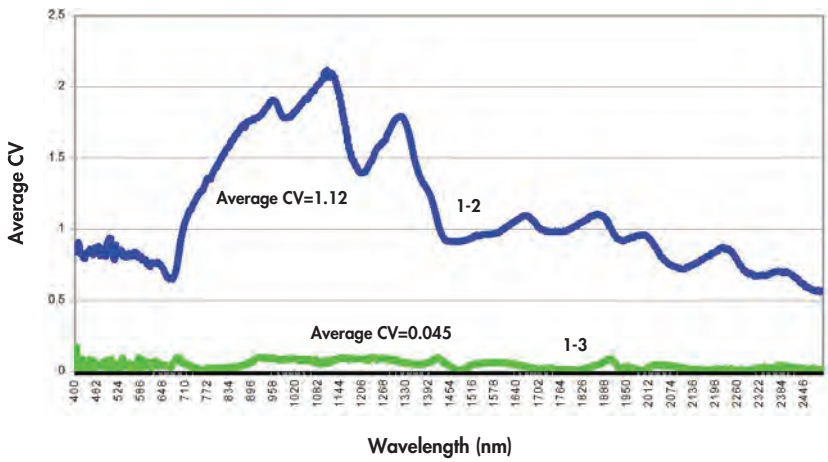


Figure 11.5 Influence of efficient mixing on the standard deviation (SD) of spectral data of a feed mix. The coefficient of variance (CV) of the dry total mixed ration (TMR) has been used. Upper line = before mixing and the lower line = after mixing.

11.12 Near-infrared discriminant analysis

Near-infrared discriminant or classification analysis is a very useful method for classifying and identifying materials by type. The technique can be used with only spectral data and no reference data are required. Figure 11.6 shows two classes of Canadian hard red wheat, CWES and CPSR classes, commercial deliveries of which are difficult to distinguish from one another visually when they occur as bulk deliveries. The Soft Independent Modeling of Class Analogy (SIMCA) system separates different classes of materials on the basis of their spectral characteristics, which, in turn, are a function of their chemical and physical make-up. In the example of these two wheat classes SIMCA separates them clearly (Figure 11.6).

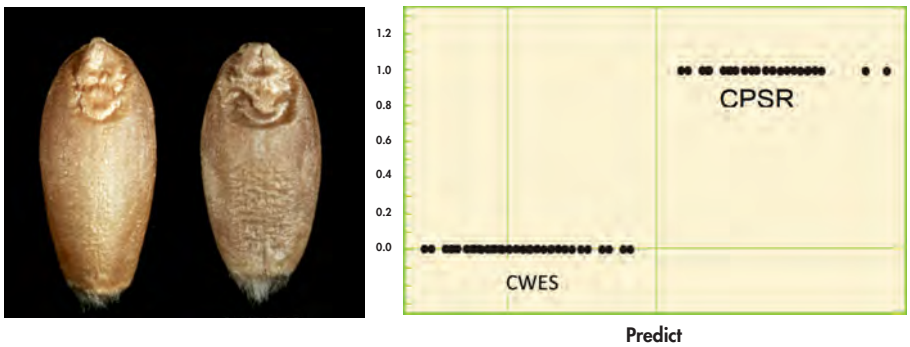


Figure 11.6 Kernels of CPSR and CWES wheat, and SIMCA classification of these wheats.

Several commercial data analysis software packages offer forms of discriminant analysis. Of these, probably the simplest is the WinISI version. Figure 11.7 shows three classes of Canadian hard red wheat, with their separation by WinISI classification. Calibration models can be developed for the prediction of different classes. These, in combination with calibrations for protein and moisture, could provide a realistic basis for electronic classification and testing of wheat at delivery.

Discriminant or classification analysis has been applied in several areas of agricultural analysis, including materials as diverse as flax straw and wheat. It can be applied to a wide variety of other materials, for example to identifying samples of fruit, meat, and other fresh products on the basis of accepted quality factors. Assembly of samples for classification analysis follows the pattern of sample assembly for composition analysis in that the variables that are likely to affect the spectra must be identified, and samples gathered that represent all of the variables – replicated several times.

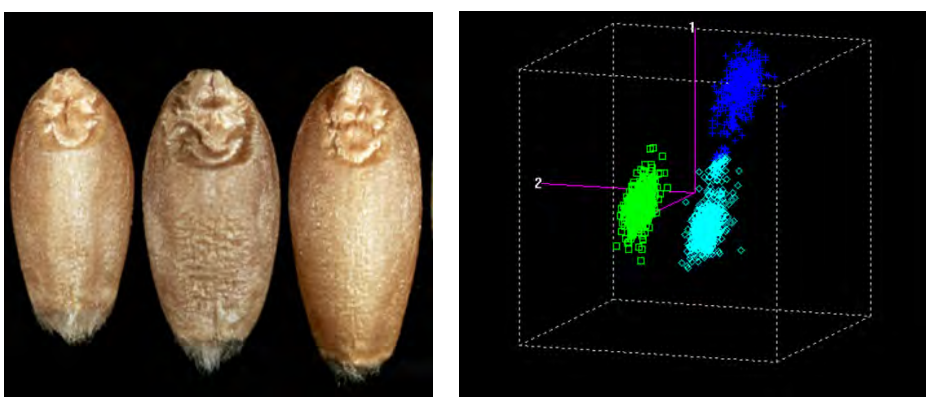


Figure 11.7 Three classes of red wheat, CWRS, CWES and CPSR with classification of these three wheats illustrated with WinISI software.

11.13 Near-infrared networking

Most modern large-scale grain handling operations that involve testing of grains and seeds for moisture, protein, oil and starch contents where a number of NIRS instruments are in use, operate all of their instruments by the use of a network. All of the instruments are connected to a 'master' instrument, located at the central laboratory that is responsible for controlling the entire operation. All of the instruments must be of the same make and model and use the same calibrations. The network is controlled from a single computer at the central laboratory. Accuracy and precision are monitored and controlled using check samples of the appropriate commodity. All of the 'satellite' or 'slave' instruments, located in areas outside of the central laboratory are interfaced to the master instrument, which is the only instrument that is calibrated and monitored by the reference methods. Accuracy is monitored by scanning check samples and all adjustments are made on the basis of results of the master instrument. The concept is discussed in greater detail in Chapter 9, section 9.2.

11.14 Peatland environmental work

A high proportion of the world's carbon is tied up in peatlands (Figure 11.8). With the progress of global warming there is some danger of these drying out to the extent that they could catch fire. Peat fires can burn for years and convert this bound carbon to CO_2. Research is in process to monitor the moisture content of peat-lands, using NIRS on location. At present the average moisture content is well above 80%, but this could change quickly to dangerously lower levels in different locations. About 60% of Canada's reserves exist on first nation reservations. NIRS is also being used to determine carbon deposits for which financial credits are awarded.

Figure 11.8 Canadian fen-type (left) and bog-type (right) peatlands.

11.15 Plant breeding

This section will be confined to wheat breeding, mainly because selection of wheat lines with the best end-use potential is rather more complicated than breeding other crops. Barley specialists may argue that point, but a) there are more different classes of wheat than there are of barley and b) the same principles apply to barley breeding anyway.

'Quality' in grain means the end-use potential of the grain. The most important factors that affect wheat quality are kernel texture, kernel colour, protein content and gluten 'strength'. Selection for quality characteristics must be carried out in the laboratory where the breeder recruits the expertise of cereal chemists to advise on the quality factors characteristic of each genotype. Screening for quality is much more time-consuming and expensive than screening for agronomic traits and disease resistance, but equally important. If the quality is not maintained in a new variety an increase of e.g. 3% in yield (significant to a plant breeder) may lead to an increase in wastage of 3–4%, which cancels the advantage. This is non-productive and damaging to the reputation of the breeder and the institute from which the new variety was released for commercial growing. It can also have economic consequences to users of the variety.

Laboratory analytical methods are a valuable asset to plant breeding. An on-going source of frustration between the breeder and the cereal chemist is that the breeder would like full analytical data, including physicochemical data and even baking data as early as the F3 generation, but cannot spare more than a few grams to the chemist because of the need to plant the selected lines. Practically any laboratory analytical process involves grinding the sample, which precludes planting the seed. Recent advances in NIRS prediction of functional parameters in whole kernels have paved the way for selection of the best (and the worst) material in early generations. Calibration models can be developed with as little as 30 g of whole kernels and the testing is non-destructive, so that the seed can be planted after testing and selection. As well, the nature of NIRS testing allows the simultaneous screening of 200–300 samples for several parameters in a normal work day; far beyond the volume of testing that can be achieved by 'traditional' methods, and at a fraction of the cost. Thresholds for acceptance or rejection based on selected characteristics could be set up using NIR classification analysis.

Wheat breeders' trials of advanced lines are usually grown in several locations and over several seasons. The volume of samples generated can place heavy stress on chemical screening due to the time and expense, but NIRS testing of whole kernels can revolutionise screening in both early and advanced generations.

Testing of whole kernels also saves the lengthy, laborious and destructive chore of grinding thousands of samples. Prior to NIRS, testing lines of a given wheat class that had been grown on several locations were usually composited before full laboratory testing, because time and expense precluded testing all the individual lines. This obscured most of the variance induced by growing location, and the very important interactions between genetics and growing location and season remained unexplored. But NIRS enables testing of all lines grown on all locations. This enables the determination of the effects of genetics/season/location interactions within a season, which provides much more information to the breeder. Lines that show excessive variability in kernel texture, gluten strength and other functional parameters should be discarded. Wheat quality parameters for which NIRS models have been developed are summarised in Table 11.5.

Table 11.5 Applications of NIRS in wheat breeding – whole grain applications

Quality factor	r^2	SEP	Mean	SD	RPD	Potential for use
Kernel colour (L*, b*)	0.96	0.30	17.4	1.1	3.7	Excellent
Kernel texture (PSI)	0.84	1.98	56.3	5.1	2.6	Fair
Farinograph water absorption	0.91	1.09	63.4	5.2	4.8	Very good
Farinograph development time	0.62	1.2	5.0	2.7	2.2	Poor
Farrinograph stability time	0.70	1.9	7.8	3.5	1.8	Poor
Farinograph mixing tolerance	0.72	17.7	52.4.	46.2	2.6	Fair
Extensigraph maximum height	0.72	85	506	228	2.7	Fair
CSP mixing energy	0.76	2.5	10.2	6.2	2.5	Fair
Protein content (%)	0.98	0.20	12.7	2.1	10.5	Excellent
HVK (%) (Durum wheat)	0.78	5.3	78.5	12.6	2.4	Fair
Barley TME	0.85	0.26	14.0	0.66	2.5	Fair
Malt fine grind extract	0.52	1.00	73.1	2.2	2.2	Poor
Oat groat (%)	0.81	0.95	74.2	2.7	2.9	Fair

CSP = Canadian Short Process bread making; TME: True metabolisable energy

To summarise, the use of NIRS testing in wheat or any other plant breeding enables screening of whole kernels and seeds in early generations, and increases the numbers of samples that can be screened by several hundred percent. It also enables the testing of large trials of advanced lines grown on different locations within a region. This provides important information on the complex genotype/location/season interactions, which was not practicable previously. The NIR technology has great potential as a screening method for quality factors in generations as early as F3. It can enable screening of very large numbers and can greatly reduce costs.

Calibrations for constituents or functional parameters, such as digestibility that call for expensive reference analysis can be developed by starting on a small scale with as few as 30 samples. The calibration model is used to predict new samples. New samples that add variance in the spectral and/or reference data can be sent for analysis, the results of which can be used to evaluate the effectiveness of the model. This procedure is continued until a large database with 200 or more samples is built up, and a stable calibration obtained. An extra advantage of this approach is that samples are likely to be assembled over a considerable time period, which will add more variance in the form of seasonal changes in the material. Setting up calibrations in this way with small sample sets is supported by use of a repeatability file (see Chapter 9, section 9.1.8).

11.16 Precision agriculture

Precision agriculture is a fairly new area of research. The objective is to map the fertility of a field to improve the efficiency of fertilizer use by applying the fertilizer more heavily where the crop has removed more of the nutrients, and less where less of the nutrients have been removed. The concept depends on the fact that the composition of plants growing in a field is a good indication of the fertility level. This principle has been successfully applied in New South Wales, Australia, where analysis of rice at a very early stage of growth is being used to provide fertilizer recommendations to farmers.

The concept was carried a stage further by workers at Montana State University where samples were taken on a combine harvester and used to monitor protein content of the wheat as it was being harvested. The harvester was equipped with a Global Positioning System (GPS) and a load cell to monitor both crop yield and protein content over the field. The data were computed into nitrogen uptake, and removal from the soil, and were used in the subsequent season to regulate the amount of nitrogenous fertilizer applied. An amount equivalent to replace the

nitrogen removed was added, together with sufficient extra fertilizer to provide an improvement in yield and/or protein content. Farmers that use the system have reported savings in fertilizer expenses of up to 20%. An equally important benefit is that farmers can identify grain of higher or lower protein content and thereby gain the maximum profit from their crops. Precision agriculture has many potential applications in both developed and developing countries.

An NIR instrument for use in this type of action has to be capable of very fast testing. Developments in Europe (called 'on-site' NIRS analysis) have enabled prediction of moisture content of fresh and haying forages during harvest, using a modified Zeiss Corona diode array NIRS instrument, which is fitted onto the forage harvester (see Figure 11.1). This instrument scans the wavelength range from 900–1500 nm using an InGaAs detector. The most difficult aspect of this innovative approach, namely providing consistent sample access to the instrument, has been overcome. The main use of on-site NIRS testing in Europe has been in the bio-fuels industry, but it has an even higher potential in sewage and manure management and precision agriculture.

11.17 Soil, peat and compost analysis

The most labour-intensive aspect of soil analysis is the sampling. Soil analysis shares this with forage analysis, but differs in that soil must be sampled from at least two depths to obtain a reasonable estimate of the nutrients present. Application of NIRS to soil analysis is further complicated by the fact that there are few absorbers in soil and most of the volume of soil is inorganic material with few NIR absorbers.

A third complication with soil analysis, whether by NIRS or reference methods, is that of sample presentation. Soils vary widely from heavy intractable clays to light sandy soils. To obtain a true picture of the soil as a growing medium, the soil should be analysed for nutrients as it is in the field. But accurate reference analysis is not possible without size reduction, and this in turn calls for drying. Drying of soil is usually carried out at 40°C or room temperature. Even at low moisture content, heavy clay soils are difficult to reduce to a condition that favours precise analysis. The best way to overcome this is by analysing several replicates. This applies to both reference 'wet chemistry' and NIRS analysis. This is much faster and cheaper by NIRS than by reference methods. Prediction of total and organic nitrogen and carbon, pH and cation exchange capacity have been the most successful applications of NIRS to soil analysis to date. Prediction of phosphorus in soils by NIRS has not been so successful to date.

Figure 11.9 shows spectra of three air-dried soils. The main effect is that of particle size. Absorption bands in the 1400 and 1900 nm areas due to water appear very prominent, even in oven-dried soil. The peak at 2206 nm has been attributed to an interaction between silicates and water. The peaks at 1400 and 1900 nm are decreased by drying, but not the peak at 2206 nm, which supports the silicate hypothesis. Figure 11.9 also presents the 2nd derivative (2 4 4 1) spectra of the same samples. Weak bands at 1978 and 2176 nm indicate the presence of nitrogen-containing constituents.

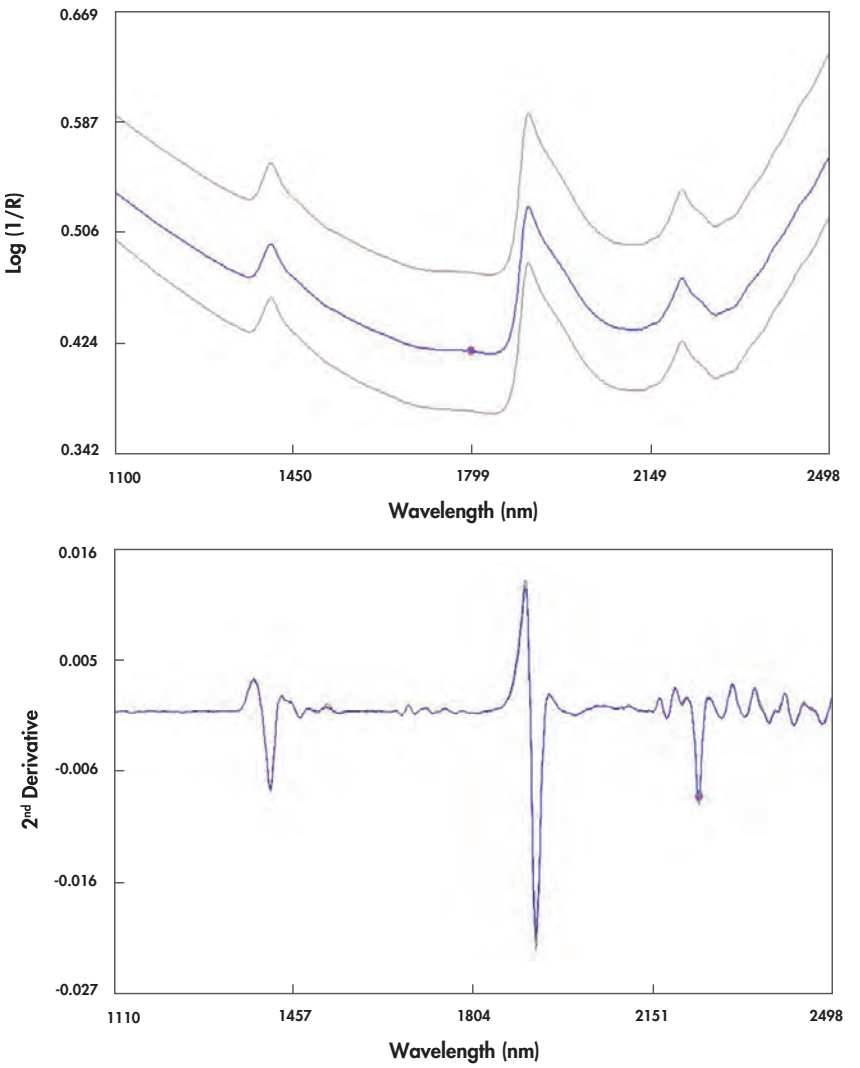


Figure 11.9 Log 1/R and 2nd derivative (2 4 4 1) NIR spectra of three dried soils.

Soil should be scanned in a form that is as near as possible to the form in which it lay in the field. Sample preparation for reference analysis of soils includes air-drying and size reduction by grinding. The samples bear little resemblance to the original soil in terms of its texture or moisture content. Soil is sampled using probes or augers that are pushed into the soil to the required depth and withdrawn, holding the soil, but ideally the soil should be scanned either in the field or as soon as possible after sampling in order to preserve the original moisture content. The ideal way of analysis of soil would be to insert a NIRS probe directly into the hole made by such augers. Two novel methods of applying NIRS to soil analysis have recently been introduced. The first uses a NIR sensor attached to a disc plough shank, and the second uses a NIR sensor attached to an augur-type probe, which can be forced into the soil to a depth of up to one metre (Figure 11.10) Soil parameters that have been measured include moisture and organic carbon contents and pH.

Figure 11.10 The Veris NIRS soil probe

Figure 11.11 show log 1/R and 2nd derivative (2 4 4 1) spectra of peat moss, which is a form of soil, and a highly organic material. There was a noticeable band shift in the 1400 nm area as a result of drying. The prominent bands at 1726, 2308 and 2348 nm are traditionally oil bands. As they are attributed to C-H stretching and deformation they could belong to constituents other than oil. For example, proteins, cellulose and starch all contain C-H groupings. The band at 2270 nm is characteristic of cellulose.

Peat itself is a preliminary to the formation of coal. Its depth ranges down to several metres below the surface and is characterised by moisture contents that range up to over 90%. Traditionally peat has been dug, dried in the air and used

as fuel. Peat moss used in horticulture occurs at or very close to the surface and is much lower in moisture content.

An aspect of soil analysis that is rapidly receiving more and more attention is that of carbon sequestration in soil. Because of the increase in greenhouse gases in the atmosphere, agricultural practices that transfer carbon from the air into the soil are favourable to the environment. Because of this interest, determination of carbon in soil has received a lot of focus. The top 10–15 cm of soil is rich in roots which are not soil. Roots should theoretically be removed before analysis of the soil by the reference method or NIRS. When NIRS is used to predict the organic carbon content of the soil, the roots (an important component of the soil) contribute a high proportion of the carbon. The form in which the carbon is stored in the soil is also important in determining how long the carbon will stay in the soil.

Calibrations have been developed for the determination of organic and total carbon in soil by NIRS. These will become valuable in future monitoring of carbon exchanges in soils. Nevertheless, more work is required to develop calibrations for various carbon fractions that turn over at different rates, particularly those that are resistant to breakdown and will remain in soil for long periods. Another type of agricultural material that is also receiving a lot of attention is compost. Composting is regarded as a practical and environmentally acceptable system for disposal of manure from cattle feed lots. Figure 11.12 show spectra of typical cattle manure compost before and after air-drying. Bands at 2308, 2348 and 2384 nm are also prominent in compost. The apparent wealth of compost in nutrients, based on absorbers in the traditional protein and oil areas, underlines the value of compost as a soil nutrient.

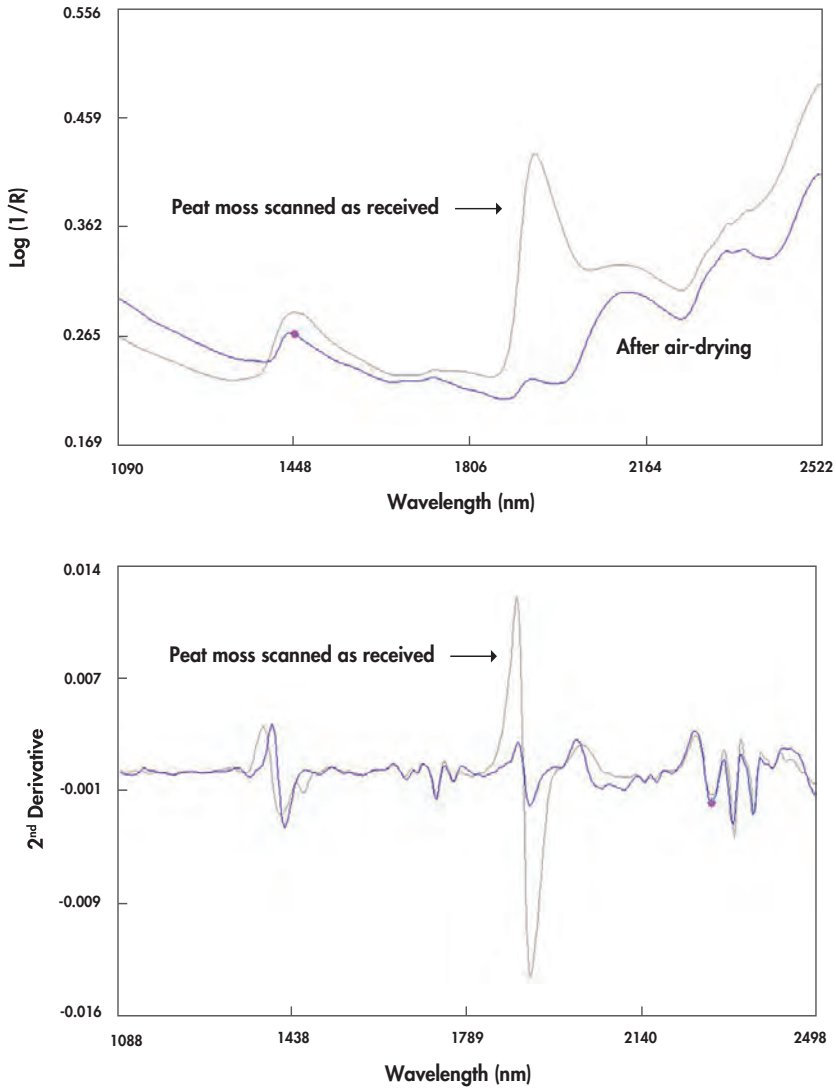


Figure 11.11 Log 1/R and 2nd derivative (2 4 4 1) spectra NIR spectra of peat moss as received and after air-drying.

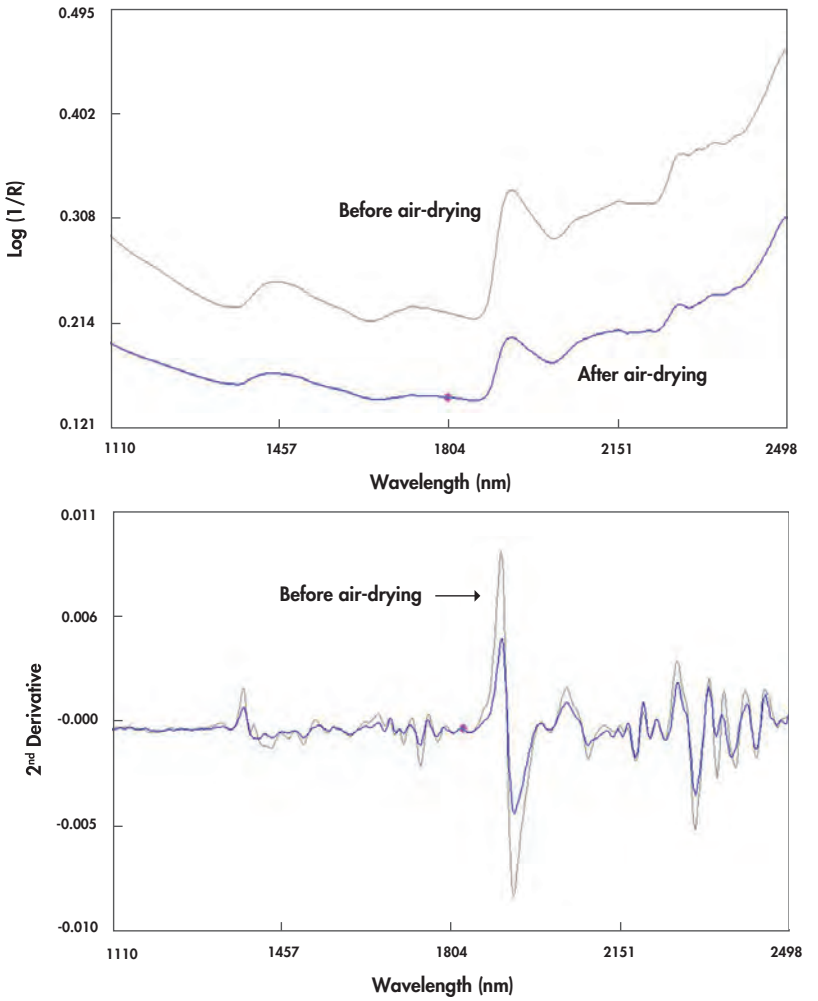


Figure 11.12 Log 1/R and 2nd derivative (2 4 4 1) spectra NIR spectra of beef cattle compost before and after air-drying.

11.18 Wine analysis

Analysis of wine begins in the vineyard with the analysis of the growing grape. The ratio of sugar content to titratable acids is an important indication of the stage of maturity of the fruit. Sugar content is measured in the field by determination of total soluble solids (°Brix) using a refractometer. When the sugar content reaches a certain level, the grapes begin to be tested for titratable acidity. This is done in a laboratory. Analysis of the must (fermenting grapes) begins early in the fermentation stage. The most important parameters tested in wine are °Brix, alcohol, pH, titratable acidity, malic acid, volatile acidity and alpha-amino nitrogen. Other parameters that have been measured include glycosyl-glucose,

anthocyanins and some sensory factors. The most advanced work on application of NIRS in the wine industry has been carried out at the Australian Wine Research Institute in Adelaide, South Australia.

11.19 Wood analysis

Wood in the form of sawn lumber (timber) has been the major factor in the construction of buildings and furniture for centuries. A third major consumption of wood is in the production of paper, most of which is produced from wood pulp. The effectiveness of wood in construction, furnishing, or pulp-production is largely a function of the species, and the growing conditions. For construction in particular, features such as density, resistance to mechanical stress, elasticity modulus, bending strength and resistance to compression are all key factors in the quality of the wood. For paper production, kraft pulp-yield potential, cellulose, lignin and moisture contents all contribute to the quality and value of the wood. NIRS has been successfully applied to the prediction of most of these parameters. Other areas of application include the prediction of the degree of degradation due to aging and changes in surface characteristics of wood to which various coatings have been applied.

11.20 Wool analysis

The moisture content of wool can be as high as 30%. High moisture content (above 14%) can cause wool to develop mildew during shipping. Contaminants of raw wool as delivered to the scour include wool wax, suint (mainly dried sweat), soil, manure, leaves and other organic matter. Wool is tested for residual grease, ash and vegetable matter, after washing. Another parameter measured in wool by NIRS include medullation, the degree to which air has entered the wool fibres. This affects the reaction of wool to dyes. Because of its springy nature, sample presentation has always been a problem in wool analysis by NIRS. Working with a KES/WRONZ NiraSpec diode array instrument, the Wool Research Organisation of New Zealand (WRONZ) have developed a sample presentation method that allows testing of 100 g of wool, which very much improved the precision of analysis.

11.21 Last words on NIRS applications

This chapter has introduced some of the many areas in which NIRS technology can be employed with great benefits in terms of time and expense, as well as opening the door to analytical work that could not be undertaken prior to the advent of NIRS.

Probably the most complicated and consistent problem overall with all the diverse applications has been sample presentation to the instrument. An NIRS instrument can only record what it sees, and calibrations and future analysis depend mainly on the quality and veracity of the spectra. This is especially important in the case of fruits and vegetables, which differ widely in size and shape. The instrument companies strive to ensure that their products accurately record the spectra of the samples presented to the instrument. It is up to us, the operators, to prepare and present the samples properly. It is particularly important to determine the reproducibility of spectral data when NIRS is to be applied to the analysis of these very different types of fresh materials. Individual operations that choose to apply NIRS to the analysis of fruits and vegetables develop their own systems for sample presentation. The reproducibility of scanning is a key aspect of these applications.

A second factor of equal importance is the moisture content of the sample as tested and as it exists. Materials such as fresh fruits, vegetables and forages, meat and fish are high in moisture and their economic and nutritional values depend on their composition at the time that they are used. Most reference analysis calls for drying and size reduction, whereas NIRS analysis can be carried out on the fresh material. For accurate calibration, it is essential to know the composition on the basis of the actual moisture content of the fresh material at the time it was scanned, and analytical results should be reported on the basis of the original moisture content.

The final word

This book has attempted to assemble all of the practical elements that operators need to know in order to get the most benefit from NIRS technology. The technology has changed and become more sophisticated over the 44 years since it was first applied to actual analytical work in industry. The basic principles are the same as they were in the early days. The wisdom and genius of engineers, software experts and chemometricians have enabled tremendous advances to be made in instruments and software. But the time per test is still about two minutes from sample to sample, because, except for on-line testing, the sample still has to be placed in the instrument and removed after testing. For its most widespread field of application, in the agricultural, feed and food industries, NIRS is the ideal method because of its flexibility in sample presentation, speed, freedom from chemicals and economics. Because of its proven superiority in precision (reproducibility) over most laboratory reference procedures, NIRS spectra are highly reproducible, and provided that the calibrations have incorporated all the

sources of variance, NIRS results can confidently be considered as being more reliable than most reference laboratories.

In the interests of time and space, the book has focused to a certain extent on grain applications. Applications in areas such as the meat, fish, fresh fruit and vegetables, tobacco and numerous other areas have also been successful. All of them use essentially the same principles as those described for grain applications.

References

AACC. 2000a. AACC International Approved Methods of Analysis. 11th Edition. Method 08-01.01. Ash – basic method. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-08-01.01

AACC. 2000b. AACC International Approved Methods of Analysis. 11th Edition. Method 08-21.01. Prediction of ash content in wheat flour – near-infrared method. AACC International, St. Paul, MN, USA: AACC International. http://dx.doi.org/10.1094/AACCIntMethod-08-21.01.

Malley, D.F. & Williams, P.C. 1997. Use of near-infrared reflectance spectroscopy in prediction of heavy metals in freshwater sediment by their association with organic matter. *Environmental Science and Technology*, 31, 3461-3467. https://doi.org/10.1021/es970214p

Manley, M. (2014) Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chemical Society Reviews,* 43, 8200-8214 (Tutorial Review). https://doi.org/10.1039/C4CS00062E

Mathias, J.A., Williams, P.C. & Sobering, D.C. 1987. Determination of lipid and protein in freshwater fish using near-infrared reflectance spectroscopy. *Aquaculture*, 61, 303-311. https://doi.org/10.1016/0044-8486(87)90158-X

Suggestions for further reading

Burns, D.A. & Ciurczak, E.W. 2007. *Handbook of Near-infrared Analysis*. 3rd Edition. New York, USA: Marcel Dekker, Inc.

Roberts, C.A., Workman Jr., J. & Reeves III, J.B. 2004. Near-Infrared Spectroscopy in Agriculture. Madison, Wisconson, USA: American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.

Index

A

B

C

D

E

F

G

H

I

K

L

M

N

T

V

W