PROJECT REPORT No. 310

THE DEVELOPMENT OF NEAR INFRARED (NIR) SPECTROSCOPY CALIBRATIONS FOR THE PREDICTION OF WHEAT AND FLOUR QUALITY

AUGUST 2003

Price: £7.50

PROJECT REPORT No. 310

THE DEVELOPMENT OF NEAR INFRARED (NIR) SPECTROSCOPY CALIBRATIONS FOR THE PREDICTION OF WHEAT AND FLOUR QUALITY

by

S J MILLAR

CCFRA, Station Road, Chipping Campden, Gloucestershire GL55 6LD

This is the final report of a 42-month project that started in January 2000. The research was funded by a grant of £189,229 from HGCA (Project No 2282).

Table of contents

1 Abstract

The protein component of wheat and wheat flour is widely recognised as having unique properties that result in the characteristic textures of the wide range of baked products produced throughout the world. A wide range of techniques have been applied to the study of these characteristics and recently a number of approaches using the size-exclusion high performance liquid chromatography (SE-HPLC) technique have been reported. One method in particular, Profilblé®, developed in France, has been applied to the analysis of UK and French-grown wheats in work reported here.

Near infrared (NIR) spectroscopy is widely used in the global cereals industry for the assessment of compositional characteristics such as protein, moisture and lipid contents. As the use of more sophisticated instrumentation and data processing approaches has become more widespread, however, so too has the development of new, more advanced applications. This study was undertaken to develop NIR calibrations for a number of measures of wheat and flour quality in general and protein quality in particular.

New NIR calibrations have been developed using a research grade scanning instrument for over 300 samples of wheat from three harvest years (1999-2001) grown in both the UK and France. Calibrations for compositional parameters gave very high levels of performance as would be expected. Calibrations for both test baking parameters and measures of protein quality, particularly those derived from SE-HPLC analysis, also indicated potential for routine use. When the calibrations were assessed using a separate validation set, the performance of those for protein quality attributes were consistent with the results for the calibration set, while those for test baking parameters were poor, indicating limited potential for routine use. Calibrations developed using a dataset combining samples analysed in both the UK and France exhibited similar levels of accuracy to those from UK-analysed samples alone, although they promised to be more robust in routine use.

Overall, UK-grown samples tended to have a balance of gluten properties in which the proportion of gliadin material was greater than those from France. The ranking of performance for varieties analysed in both locations using different test baking regimes was very similar, however, the differences in gluten properties observed notwithstanding. There was some evidence, however, that the performance of flour in the French breadmaking test was not so reliant on higher levels of protein content as was generally the case for Chorleywood Bread Process assessment in the UK.

The new calibrations are now commercially available for users and further information about them may be obtained from the author (s.millar@campden.co.uk).

2 Summary

Introduction

Wheat is generally regarded as being unique amongst cereals as a result of the properties of the storage proteins laid down during grain development. These proteins are the major contributors to the ability of doughs produced from wheat flour to retain gas produced by leavening agents and thus to produce the characteristic texture sought after in a wide range of baked goods. All plant proteins may be separated into a number of different fractions which differ in solubility, but those from wheat include a proportion of insoluble material. This is thought to comprise the polymeric proteins, dominated by the high molecular weight glutenin fraction. A range of techniques have been applied to the study of this group of proteins and, of these, size-exclusion high performance liquid chromatography (SE-HPLC) has been one of the most informative, particularly in characterising the effect of protein molecular weight distribution on baking potential. The most recent developments by a team made up of the former Institut Technique des Céréales et des Fourrages (ITCF, now part of ARVALIS) and l'Institut National de la Recherche Agronomique (INRA) have led to the development of the technique to a stage at which separation of high molecular weight glutenins, low molecular weight glutenins, gliadins and a combined albumin and globulin fraction may be carried out. This method, Profilblé®, has been used within this study. In addition to the characteristics which result in the functionality collectively called 'protein quality', the amount of protein within the grain, flour and ultimately dough also contributes to the baking performance achieved and it is this which has resulted in the routine assessment of wheat and flour protein content across the world.

Near infrared (NIR) spectroscopy is the most widely-used technique for measuring the composition of cereals and their products and, in particular, their protein content. Until relatively recently, however, the use of the technique for the assessment of characteristics related to functionality has been more limited. There have been a small number of studies in the last few years, however, which demonstrate potential for research grade instruments to be used for this purpose with wheat and wheat flour. Early work concluded that the performance of an NIR calibration for wheat SDS sedimentation volume was no more predictive than simply using protein content alone. This demonstrated that the protein content effect on the development of calibrations for protein quality was a significant factor which needed to be recognised. More recently, however, the removal of the effect of protein content on protein quality calibration performance has been demonstrated, indicating the potential of developing NIR calibrations for rapid assessment of wheat protein quality.

Although this potential is clear, until now no suitable NIR calibrations have been developed that may be used by breeders, producers and processors in the UK. Given the wide range of material available through the annual harvest surveys undertaken in both the UK and France, as well as the expertise in NIR at CCFRA and ARVALIS, discussions with representatives from HGCA led to the project reported here. The overall aim of

the project was to develop new NIR calibrations which could be used by the cereals communities in the UK and France and within this, 3 specific objectives were identified:

- To assess the application of size-exclusion high performance liquid chromatography (SE-HPLC) for quality testing of UK and French-grown wheat samples.
- To develop accurate and robust NIR calibrations for wheat and flour quality indices generally and protein quality characteristics, in particular, as measured using SE-HPLC.
- By undertaking a programme of sample exchange to highlight the performance of UK and French-grown wheats under alternative milling, quality testing and baking regimes.

Materials and Methods

More than 300 samples of UK and French-grown wheats from 3 harvest years (1999-2001) have been milled, analysed and test-baked [Chorleywood Bread Process (CBP) and no-time dough (spiral mixer) methods] in accordance with usual practice for the annual UK harvest survey. In addition, SE-HPLC characterisation of the main protein fractions has been carried out. NIR spectra were collected from the whole wheat, ground wheat (as is typically used at mill intake) and for Bühler milled white flour. Near infrared calibrations were developed for each of the parameters of interest for each of these materials. A range of different data processing techniques were evaluated and the best were used to develop calibrations using the entire dataset. In addition, a proportion of the set was removed during calibration and used as a separate validation set in order to test the likely performance of the calibrations under typical 'real-world' conditions.

To ensure that the calibrations developed would be robust when used for flour milled under different conditions, a combined dataset was also generated containing more than 400 samples which had been milled at CCFRA and ARVALIS. Calibrations were then developed following the protocols developed for the UK generated samples.

Results and discussion

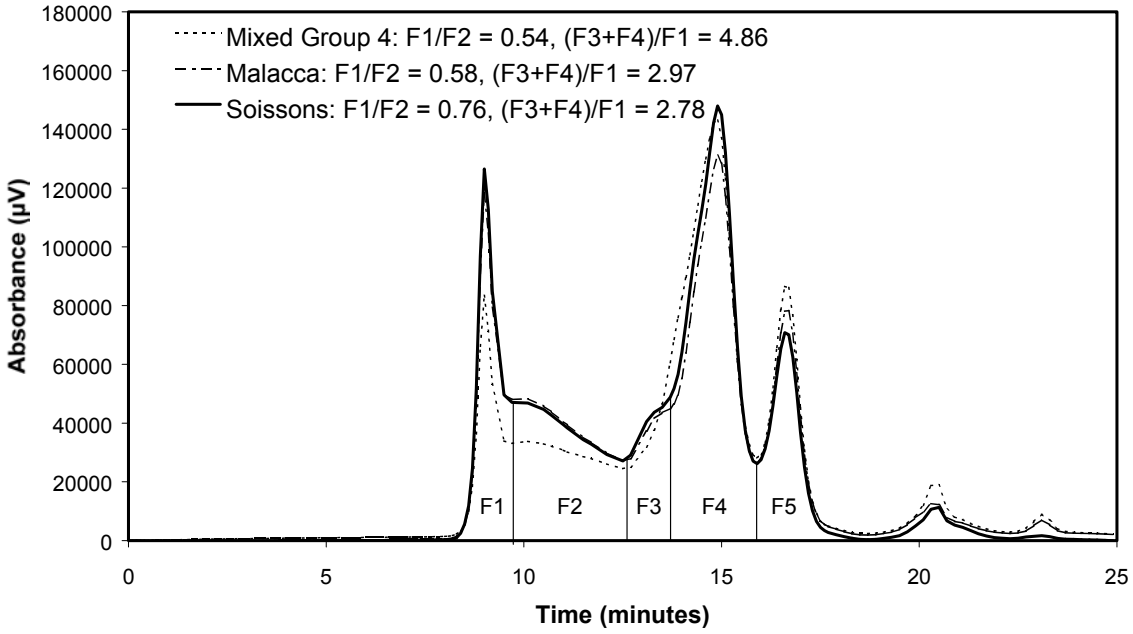
Flour SE-HPLC analysis

Chromatograms had a general form of that illustrated in Figure a. Peaks indicate time periods across which material of a particular size eluted from the column. Due to the way in which the columns used function, physically larger material eluted earlier than smaller material. The different fractions have previously been assigned to high molecular weight glutenins (F1), low molecular weight glutenins (F2), gliadins (F3 and F4) and a combined albumin and globulin fraction (F5).

An example of the information which may be generated using the technique may be derived from the relative performance of each of 3 UK-grown wheats which represent 'weak feed' (Group 4), 'standard bread'

(Malacca) and 'extra strong bread' (Soissons) gluten characteristics. These chromatograms have been normalised for protein content, allowing differences in shape to be more clearly observed. The flour from the Group 4 wheats clearly shows a significantly lower response for F1 and F2, indicating the relative lack of large, polymeric proteins. This results in the larger (F3+F4)/F1 ratio results. Although both breadmaking varieties had fairly similar results for (F3+F4)/F1, the difference in their gluten properties is apparent from the difference in F1/F2 ratio observed. That for Soissons was greater than that for Malacca, representing a shift within the glutenins toward relatively more of the higher molecular weight fraction.

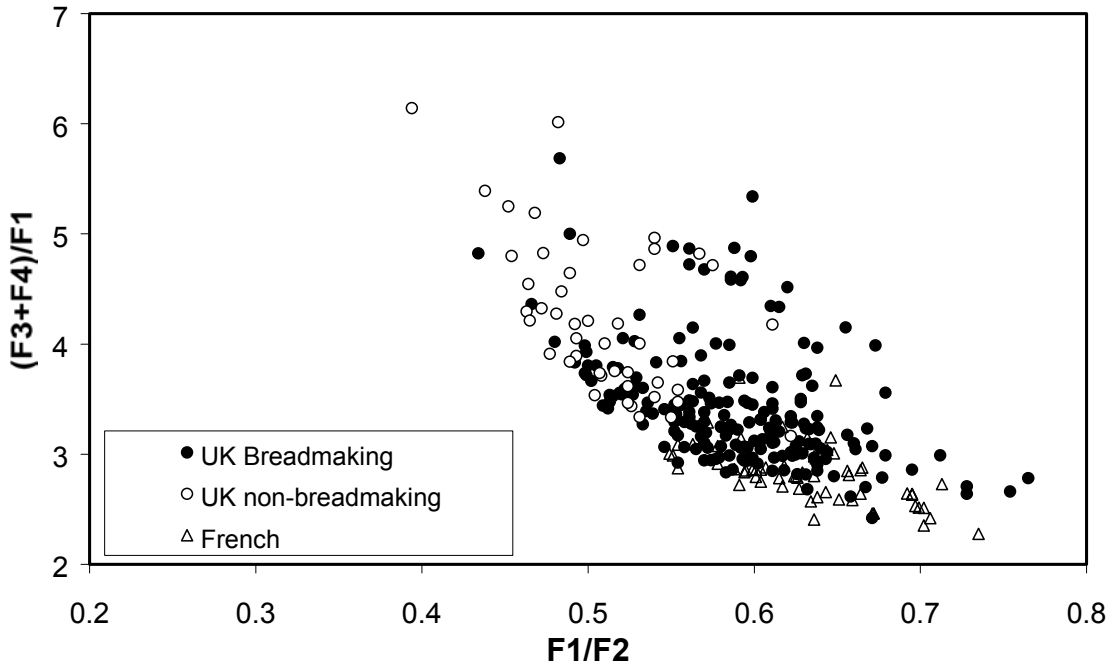
Figure a. Classification of typical UK-grown varieties using SE-HPLC results (area under chromatogram normalised for protein content)



On the basis of this limited comparison, it seems that the (F3+F4)/F1 ratio classifies flours broadly into breadmaking and non-breadmaking classes, while the F1/F2 ratio may be used to discriminate 'extra strong bread' from 'standard bread' categories. To understand whether this is consistent for a range of different sample types, the entire dataset as well as the main categories represented within it were grouped using these two parameters (Figure b). The first observation arising from this work is the underlying relationship between F1/F2 and (F3+F4)/F1 for the dataset collected. This is not surprising as F1 is a common variable in both ratios and it would be expected that varieties tending to have greater proportions of higher molecular weight material would tend to have higher F1/F2 values and correspondingly lower (F3+F4)/F1 results. To assess how broad categories could be discriminated, the results obtained for all the UK samples collected which were either grown and supplied as biscuit or feed varieties or which were supplied commercially as non-breadmaking were compared with those classified as having been grown for breadmaking. In addition,

the French-grown samples were classified separately. This shows that the breadmaking varieties generally tend to have higher F1/F2 and lower (F3+F4)/F1 values as would be expected given the fact that they form more elastic doughs and, therefore, that their proteins will tend to produce more extensive cross-linking and will tend to have more material in the higher molecular weight fractions. The samples do demonstrate a range of values, however, with different properties being evident from different relative proportions of higher molecular weight glutenins and gliadin material. There was also some overlap between the groups with a number of the lower grade wheats giving similar results to those obtained for breadmaking varieties. This would be expected as protein characteristics and the baking performance obtained will be based on a continuum even though several sub-classes within this will be evident.

Figure b. Classification of UK and French-grown wheats using SE-HPLC results



It is also interesting to note the tendency towards values which are thought to result in stronger gluten properties for the French-grown samples. Although the majority give similar values to those obtained for UK breadmaking varieties, a number give high F1/F2 and low (F3+F4)/F1 values. This may be a reflection of both the different genetic backgrounds of these varieties, but will also relate to growing conditions given that in some cases, the same varieties had been grown in both locations. It is also clear that the different breadmaking methods used in the two countries will also have an impact as different gluten properties will be required and, therefore, selection and development of varieties will be carried out using different criteria.

Calibration performance for the entire UK calibration database

Calibrations were developed using all the best data pre-treatments established during development of calibrations using 1999 harvest data only. In each case, cross-validation was used to allow the optimum number of terms for inclusion to be assessed and to give standard error values which were more meaningful in terms of likely future predictive potential. In addition to the squared correlation coefficient (R^2), which is the proportion of the variation in the dataset explained by the regression equation, two other error terms were used as a measure of calibration performance. The first was the standard error of calibration (SEC), which is the error associated with the calibration when tested using all the samples in the set. This tends to give an optimistic view of likely future performance. The second term is the standard error of cross validation (SECV). This is a measure of the performance of the calibration when samples are iteratively removed from the calibration set, predicted and then returned to the calibration set. Typically this gives a better estimate of likely future performance.

Generally the results for flour were better than those for ground wheat and, in turn, those for whole wheat (Table a). Although SEC and SECV may not be easily used to directly compare results for parameters having different scales, the squared correlation coefficient may be used to give a 'rule of thumb' assessment of the relative performances for each of the parameters. The results for protein content in each were excellent demonstrating high values for R^2 and low errors when judged by SEC and SECV statistics. The results for flour were lower than has been previously reported and demonstrate the high quality of the data generated throughout the study.

The results obtained for test baking parameters (CBP and no-time dough loaf volume, CBP crumb score) were generally poorer than those for the SE-HPLC measures of flour functionality, particularly for ground and whole wheat. This is because the test baking process is affected by flour protein characteristics, but also relies on other flour properties in addition to introducing variability as a result of the test baking procedure itself. As the transformation from flour to bread is not included in the calibration process, i.e. the calibration is being used to predict performance on the basis of the pre-transformed material, it is not surprising that performance is (a) limited and (b) progressively poorer the less transformed is the material assessed (whole wheat for example).

The results obtained for the SE-HPLC indicators of flour functionality demonstrate good performance which would allow them to be used for rapid screening of samples, particularly if flour was analysed. This indicates that NIR has clear potential to be used as a screening tool for protein quality determinations.

Calibrations were also developed using a combined dataset comprising samples analysed in accordance with French methods of assessment as well as those used in the UK (Table a).

Table a. Performance of NIR calibrations developed using datasets generated in the UK and France

Dataset	Material	Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
UK only	Flour	Protein content as is (%)	291	9.70	1.20	>0.99	0.05	0.05
		CBP loaf volume (ml)	286	3326.0	185.9	0.62	114.6	128.4
		CBP crumb score	292	6.1	1.1	0.54	0.7	0.8
		No-time dough loaf volume	212	1500.4	122.2	0.56	81.4	95.8
		F1/F2	303	0.584	0.058	0.64	0.035	0.040
		(F3+F4)/F1	302	3.432	0.646	0.82	0.277	0.326
	Ground wheat	Protein content as is (%)	285	10.89	1.10	>0.99	0.08	0.10
		CBP loaf volume (ml)	286	3326.6	186.1	0.40	143.8	156.7
		CBP crumb score	294	6.1	1.1	0.21	1.0	1.0
		No-time dough loaf volume	208	1505.4	121.9	0.58	78.9	86.4
		F1/F2	299	0.585	0.058	0.55	0.038	0.044
		(F3+F4)/F1	302	3.426	0.650	0.74	0.334	0.398
	Whole wheat	Protein content as is (%)	284	10.66	1.06	0.99	0.12	0.15
		CBP loaf volume (ml)	282	3332.0	185.2	0.41	142.7	148.5
		CBP crumb score	284	6.2	1.0	0.26	0.9	0.9
		No-time dough loaf volume	211	1507.4	123.7	0.37	97.9	103.9
		F1/F2	298	0.584	0.058	0.57	0.038	0.045
		(F3+F4)/F1	296	3.386	0.609	0.49	0.433	0.454
Combined	Flour	Protein content db (%)	385	11.05	1.46	>0.99	0.06	0.07
		F1/F2	420	0.588	0.057	0.53	0.039	0.042
		(F3+F4)/F1	420	3.369	0.626	0.80	0.279	0.334
	Ground wheat	Protein content db (%)	402	12.23	1.28	0.99	0.11	0.13
		F1/F2	419	0.590	0.056	0.53	0.039	0.043
		(F3+F4)/F1	413	3.339	0.585	0.70	0.321	0.372
	Whole wheat	Protein content db (%)	406	12.23	1.29	0.98	0.16	0.18
		F1/F2	417	0.589	0.056	0.59	0.036	0.041
		(F3+F4)/F1	412	3.331	0.587	0.57	0.385	0.408

[a]Standard error of calibration
[b]Standard error of cross validation

The results obtained for flour when using the combined database were all very slightly inferior to those achieved for the UK database. Although the difference in each case was small, the consistency of the direction of the differences would indicate that this was important. However, the milling procedure used at the two sites was quite different and would clearly result in flours from common wheats having different extraction rates, bran content and particle size. Rather than being a negative finding, therefore, the fact that the two databases may be combined to give results not far removed from those when using either of the sets alone, is very positive as it indicates significantly improved robustness. This would give users confidence that the calibrations would be more likely to perform well with the milling procedure that they use. In addition, the results for ground and whole wheat also demonstrated improvements in accuracy in some cases when compared with calibrations developed using the UK-analysed samples alone.

The overall high level of performance for the combined datasets is important for calibrations of this type as the subtle spectral variation which has been modelled tends to mean that the resultant calibrations will be more sensitive to differences between instruments. The results here indicate that this is a real possibility but is one that may be overcome by ensuring that calibrations are developed using datasets representing the range of variation (for all parameters) which may be encountered in the future. As such, the results should give users confidence in the likely performance of the calibrations for their own situation.

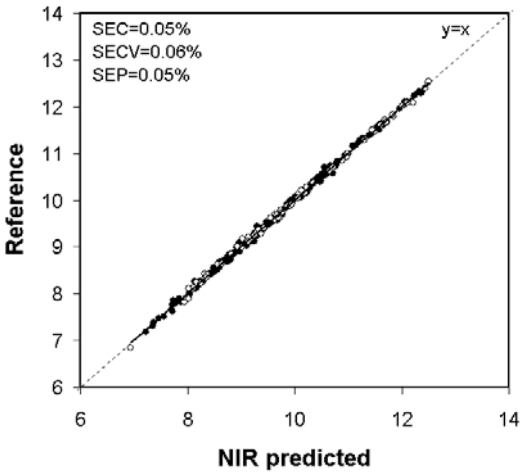
Calibration performance using separate validation samples

The UK database was split into separate calibration and validation sets to test the performance of the calibrations under conditions more typical of those where NIR is used to assess unknown samples on a routine basis. The results generally indicate that the performance of the calibrations was consistent for all the various indicators of accuracy at each stage (Figure c). In this case, representative graphs (those for flour) have been plotted to illustrate the level of performance obtained for all the materials (flour, ground wheat and whole wheat) assessed. Values for standard errors of calibration (SEC) were usually lower than those for either cross-validation (SECV) or separate validation set predictions (standard error of prediction - SEP) and this is as expected given that the SEC is the most optimistic assessment of future calibration performance. In each case the SEP was calculated as the complete error around the line of equivalence and was not corrected in any way for systematic errors (skew or bias). The regression lines in each case were plotted for the validation samples only, as the calibration samples are constrained by the regression procedure to have a slope of approximately 1 and a bias of approximately 0.

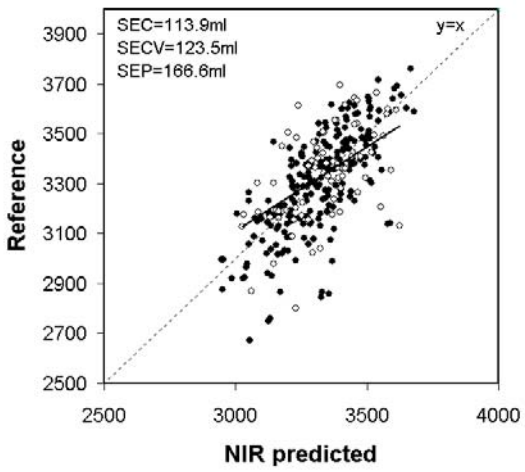
Given the range of samples collected and analysed throughout the project, it was hoped that where calibrations were developed, their performance would be sufficiently robust to accurately predict other samples not included in the calibration set.

Figure c. Performance for NIR calibrations developed using samples assessed at CCFRA

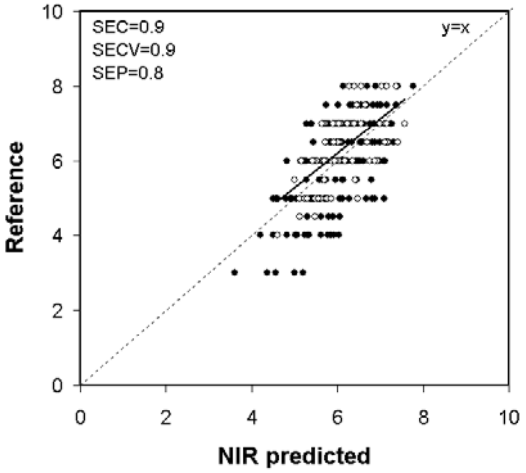
Protein content as is (%), N × 5.7



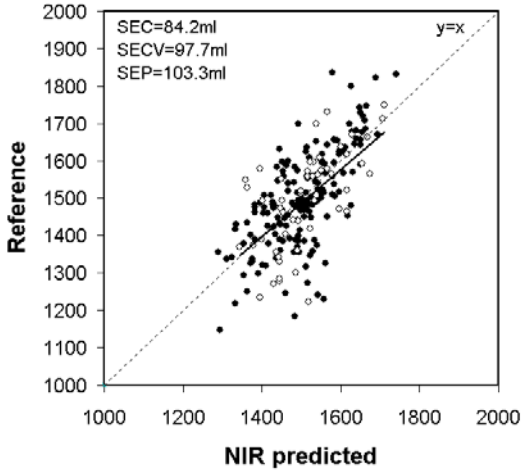
CBP loaf volume (ml)



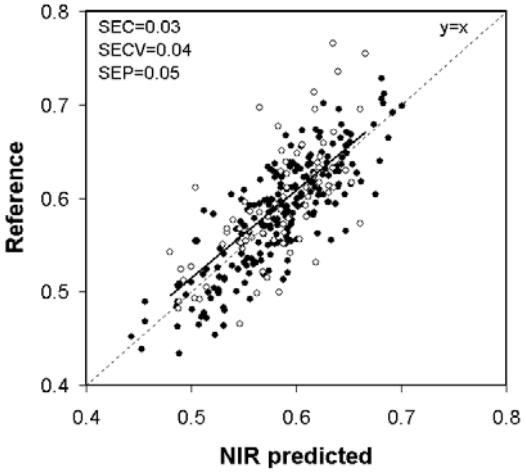
CBP crumb score



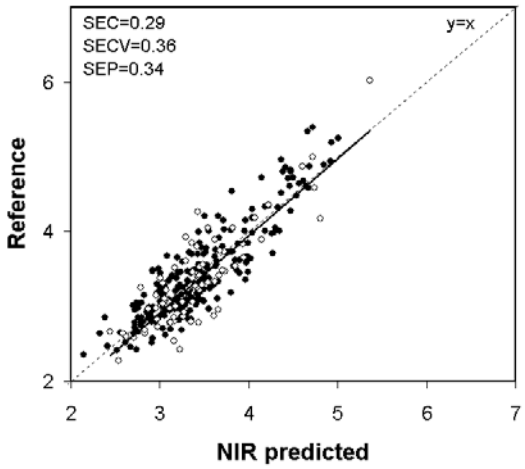
No-time dough loaf volume (ml)



SE-HPLC F1/F2



SE-HPLC (F3+F4)/F1



Where the prediction was prone to systematic error such as bias or slope deviations from 0 and 1 respectively, then the performance of that calibration should be regarded with caution.

As for the calibration development stage, the performance of protein content calibrations stands out for all the materials as being particularly good. For the parameters associated with test baking, the poor performance was due to the lack of a good underlying model. This is because of the processing involved (the effects of which will not be included in the NIR spectrum of wheat and flour) which interacts with flour quality (which may be related to the spectral data) to produce a quality which is not easy to predict even using the reference techniques.

The remaining SE-HPLC calibrations both appeared to give reasonably consistent performance with little indication of systematic errors. It is interesting to note, however, that for the parameter which is thought to relate, in particular, to the molecular weight distribution of the polymeric proteins, i.e. F1/F2, the calibration appears to consistently under-predict the samples at the top end of the range assessed. In addition, the performance of this calibration generally was lower than that for (F3+F4/F1). These observations tend to indicate that fundamentally the two techniques (the reference method and NIR) do not measure the same entity. Rather, the information in the NIR spectrum relates to another element of wheat quality which relates well to protein molecular weight distribution for much of the range but fails to model the variation exactly, particularly at the higher end of the scale of interest.

Overall, therefore, it may be concluded that the use of calibrations developed using SE-HPLC parameters will give more robust NIR calibrations than those developed using measures of test baking performance. The performance of these latter calibrations is sufficiently poor as to render them ineffective for routine use. Both of the calibrations developed for SE-HPLC parameters may be offered for use but, of these, that for (F3+F4)/F1 is thought more likely to give higher levels of performance on the basis of the calibration and validation exercises carried out during this study.

Conclusions

- Size-exclusion high performance liquid chromatography (SE-HPLC) may be used to characterise UK and French wheats in terms of their protein quality.
- NIR calibrations for protein quality and test baking characteristics may be developed which appear to have reasonable potential for rapid assessment of sample characteristics.
- Calibrations developed using a combined dataset of samples analysed in France as well as in the UK gave more robust calibrations, indicating improved potential for general use.

- Good performance was obtained for measures of protein quality when using a separate validation set while calibrations for test baking performance gave relatively poor results.
- The use of calibrations as a direct measure of test baking performance may not be recommended due to this poor predictive ability.

Implications for levy payers

Assessment of wheat functionality is one of the most important requirements of any testing regime employed by those who develop varieties, grow wheats and process the grain thus produced. It is for this reason that NIR is so widely used throughout the world for the rapid analysis of wheat protein content. While an important determinant of likely baking potential, it has long been recognised that the functionality of that protein is arguably the most important characteristic of any sample of wheat. The rapid analysis of these properties has, therefore, long been viewed as a 'holy grail' for researchers.

New NIR calibrations have been developed within this study which demonstrate good potential for use in the rapid assessment of wheat protein quality. The performance of these calibrations is generally better for flour than for ground wheat and is poorest for whole wheat. Nevertheless, useful information which could be used as a basis for screening-type assessment may be derived from each of these materials. The calibrations have been developed for use with research grade NIR instruments, which are used only in some sections of the grain industries. However, the cost of such instrumentation is now much closer to that of the instruments which are more typically found in grain receival areas, and the additional information which may be derived from their use would make their purchase feasible for many users. In addition, the availability of such instrumentation at many of the laboratories involved in the annual harvest survey means that decisions about the quality of new and established varieties of wheat may be ascertained using the calibrations developed here. From harvest 2003, the calibrations derived in this study will be used by CCFRA in the assessment of baking quality of potential Recommended List varieties carried our on behalf of Crop Evaluation Ltd.

Further information about the calibrations, their availability and use may be obtained from CCFRA, HGCA or ARVALIS.

3 Introduction

3.1 Background

Wheat 'quality' or 'fitness for purpose' is assessed on an ongoing basis throughout the wheat producing and consuming world. However, perhaps the most targeted approach is that adopted at time of harvest, where gross shifts in quality may be highlighted and where the performance of newly introduced varieties may be evaluated. In the UK, the assessment of quality of new varieties takes into account both agronomic and performance related characteristics and those that relate to milling, bread and biscuit making quality, and is carried out by a number of organisations including CCFRA.

This quality testing is based on assessment of the breadmaking and biscuit making performance of new and established wheat varieties. In making decisions, however, a broad range of quality attributes are determined. In general, these are designed to cover the three main elements of wheat quality, the relative importance of which vary with proposed end-use.

1. **Wheat endosperm texture** which determines whether the grain will be hard or soft milling. Hard milling types are generally better suited for breadmaking as the resultant flour contains a greater proportion of damaged starch which absorbs more water and is more susceptible to enzymatic attack. Soft milling wheats are sought for biscuit, cake and batter production.

2. **Level of endogeneous *alpha*-amylase activity** which determines how the various substrates will be used. A certain level is beneficial during breadmaking as a means of digesting starch to produce sugars for yeast to feed on. An excess is undesirable generally, but is particularly unwanted during bread production where the production of dextrins late in baking result in sticky and soft crumb.

3. **The correct balance of protein content and characteristics** depending on the product to be made. Bread production requires sufficient protein to be present but this protein must also have the correct rheological characteristics when hydrated to allow the gas produced during proving and baking to be retained. The characteristics required for biscuit production tend to be lower protein contents and gluten properties which result in extensible doughs.

Although these are underpinning to the majority of the wheat, flour and dough quality testing procedures currently used, the results from such studies still do not allow baking potential to be reliably predicted. A study based on samples of Swedish wheat from 2 harvest years demonstrated that although protein content could explain 47.5% of the variation in test-baked bread volume, the addition of other variables related to starch and non-starch polysaccharides properties could only increase this to 65.4% (Andersson *et al.*, 1994).

While this performance is interesting, it would be insufficient to allow consistent prediction of baking potential.

Near infrared (NIR) spectroscopy is widely-used in the cereals and related industries for the rapid analysis of wheat and flour properties. Given the wide range of quality parameters assessed during the annual harvest survey, an opportunity for the development of a broader range of calibrations was apparent with the possibility of developing calibrations which could allow rapid prediction of wheat and flour quality or baking potential. Discussion between representatives from HGCA, ARVALIS – Institut du Végétal (then Institut Technique des Céréales et des Fourrages – ITCF) and CCFRA resulted in the proposal of a joint exercise within which samples from both the French and UK harvest quality surveys would be included. It was decided that the primary aim of the work should be the development of NIR calibrations for appropriate reference techniques describing the characteristics of the protein from these samples. It was felt that by doing this, better calibrations would result compared with trying to predict baking performance directly, which is highly dependent on baking procedure used (of which there are many), and said calibrations would be more transferable to different users as a result.

3.2 The characteristics of wheat proteins

The structure and function of wheat proteins have been studied for many years and many aspects of their characteristics are well understood. However, a complete picture of their interaction and the mechanisms by which they result in the characteristic textures of baked products still eludes researchers and remains an area of active research. It is known, however, that two broad types exist, with those having a metabolic function being separated from those used as storage for grain germination. It is also known that it is this latter fraction which is responsible for the formation of gluten on hydration and that it is these, gluten-forming proteins, which are unique in wheat. Certainly, other cereal types do have some similarities but even for those that may be used in bread production, such as rye, the properties of the proteins are sufficiently different as to result in significantly lower ability to retain gas and thus to deliver the light, aerated texture of baked goods made using wheat flour.

3.2.1 Protein content

It has long been recognised that the proportion of the flour which is proteinaceous has a direct impact on its breadmaking or baking potential. Finney and Barmore (1948) were amongst the first to demonstrate a direct and effectively linear relationship between flour protein content and loaf volume which was made more direct when assessed within individual wheat varieties. This persists as an underlying quality criterion to this day with wheat protein content being one of the main characteristics assessed for trading purposes. However, the fact that each variety gave a slightly different relationship and that the gradients for different wheats (i.e. the increase in loaf volume as a function of increasing protein content) differed (Bushuk *et al.*,

1969), led to the conclusion that mere protein content is not the only factor of importance. It is for this reason that investigations into the role of so-called protein quality have become such an important part of modern cereal science.

3.2.2 Wheat protein classification

One of the earliest studies of protein structure used wheat protein as a model and through differences in solubility and functional properties, four main classes were identified (Osborne, 1907) which still have relevance in our understanding to this day (Table 1). It was some years before this information was used in an applied way to understand wheat protein function in baking, but by the end of the 1960s a programme of work at both the Canadian Grain Commission and The University of Manitoba culminated in the first publication demonstrating the existence of 0.05M acetic acid soluble and insoluble glutenins (Chen and Bushuk, 1970). This was soon followed by a publication which demonstrated very clearly the positive effect that increasing the level of insoluble glutenin had on baking performance (Orth and Bushuk, 1972).

Table 1. Osborne fractions of plants generally and wheat proteins specifically

Solubility	Plant protein fraction	Wheat protein fraction
Water soluble	Albumins	Albumins
Saline soluble	Globulins	Globulins
Aqueous alcohol soluble	Prolamins	Gliadins
Remaining insoluble fraction	Glutelins	Glutenins

Subsequent work led to the development of the current understanding of the effect of varying gliadin and glutenin proportions in wheats. The former are thought to confer extensibility to doughs while the elastic component arises from the glutenin fraction. The importance of the glutenin fraction was further demonstrated by Graveland *et al.* (1979, 1985) by developing a separation technique producing a gel protein fraction. Subsequently, the rheological properties of this fraction were used to further characterise wheat proteins having different functional properties (Oliver and Pritchard, 1993). These and other studies have now resulted in an understanding broadly based on polymer theory of how large molecules interact which indicates that for elastic properties long range order and very large molecular weights (Southan and MacRitchie, 1999) are required. Large molecules in this context make analysis difficult as it is their very insolubility which is thought to give them their functional attributes and which also makes their extraction in unmodified form impossible.

The technique which has perhaps been most widely used for analysis of wheat protein molecular weight distribution is size-exclusion high performance liquid chromatography (SE-HPLC). The traditional approach has been to use sodium dodecyl sulphate (SDS) detergent as a means of extracting the majority of the

proteins and to quantify the remaining insoluble fraction by mass (Dachkevitch and Autran, 1989). More recently the use of sonication has been proposed as an effective tool to study the structure of wheat proteins (Singh and MacRitchie, 2001) and this is an important aspect of the method used in this study (Morel *et al.*, 2000). In this, the use of a sonication step is highly controlled such that virtually all of the proteins are extracted in a form that allows both high and low molecular weight glutenins to be resolved on the column.

3.2.3 Molecular structure

Early workers trying to develop models linking gluten structure to function recognised the importance of the cysteine residues contained in the glutenin fraction. It was reasoned that the structure of gluten had to be extensively cross-linked to produce the resistance to extension typical of wheat flour doughs; the formation of disulphide bonds between and across individual glutenin units was proposed as the means by which this occurred (Schofield, 1986). These disulphide bonds were thought to emanate from the thiol groups present on the cysteine residues in gluten proteins. Evidence for this was supplied by the fact that oxygen was shown to be a key requirement for dough development and that many dough improvers act as oxidising agents.

More recently, however, the existence of dityrosine bonds within gluten structure has been proposed (Tilley *et al.*, 2001), particularly during the later stages of processing, and further work in this area will undoubtedly re-examine the accepted theories relating to disulphide bond formation.

The use of mid infrared spectroscopy has also been proposed for the study of gluten structure, although this has tended to be in the context of the hydrated system as would be found in dough (Belton, 1999; Shewry *et al.*, 2001). Nevertheless, the findings of these studies indicate the importance of hydrogen bonding in gluten structure development, both in the developing grain and during dough mixing, and clearly have an increasingly important role to play in our understanding of these systems.

3.2.4 Genetic control of gluten function

Although this aspect of gluten structure and its control is outside the capability of NIR for direct measurement, work in this area has been one of the most significant developments from the last 20 years in cereal science. Work by Peter Payne and his colleagues at Plant Breeding International (Payne *et al.*, 1987)) showed that there were particular combinations of the glutenin subunits expressed on the 1A, 1B and 1D chromosomes that gave the best potential for good breadmaking. Scores were assigned to all the different subunit combinations and the proportion of the variation in baking performance was estimated to be of the order of 47-60%. This work resulted in very significant interest in this area, with many groups subsequently studying this in great detail.

3.2.5 Use of NIR for prediction of wheat protein quality-related parameters

Although the use of NIR for the assessment of cereal composition is widespread, there have been relatively few studies aimed at the development of calibrations for wheat protein quality or related parameters. It is likely that the difficulty of using even the reference methods to effect such a judgement is one of the reasons for this. Nevertheless, a number of workers have attempted to develop suitable calibrations, with variable results. The first reported study of this kind was that of Osborne (1984), where it was concluded that the performance of an NIR calibration for wheat SDS sedimentation volume was no more predictive than simply using protein content alone. This demonstrated that the protein content effect on the development of calibrations for protein quality was a significant factor which needed to be recognised.

The next reported study assessed the performance of calibrations for a number of flour functional characteristics including two measures of baking performance, loaf height and crumb visual texture scores (Delwiche and Weaver, 1994). Here the performance of calibrations for water absorption was described as moderate, but that for all the dough and bread quality related parameters was poor, particularly for validation sets. This indicated the importance of testing such calibrations prior to implementing them as in many cases the calibration statistics were flattering and not representative of the predictive performance of the models developed.

Calibrations have also been developed using spectra taken from whole wheat kernels (Pawlinsky and Williams, 1998). This study was the first to show some potential although the performance of a large number of calibrations was reported in the work and consequently the assessment of the performance of each of the calibrations was less detailed. This paper was interesting, however, in that the use of wavelengths associated with lipid were proposed as being of importance in the development of calibrations related to protein functionality.

Delwiche *et al.* (1998) were the first to remove the effect of protein content on protein quality calibration performance, demonstrating that some underlying relationship could be found. In addition to measures used previously, such as sedimentation type methods and dough rheological characteristics, this study was the first to produce NIR calibrations based on SE-HPLC assessment of wheat protein quality. Although their performance was significantly inferior to those for protein content, these calibrations were concluded to be of potential use for screening type applications.

More recently, a different calibration approach has been taken in which the spectra from materials representing wheat starch, glutenin and gliadin material have been used to model the proportion of each material in unknown samples using a curve-fitting approach (Wesley *et al.*, 1999). Again, the importance of lipid wavelengths in the models for glutenin content was demonstrated. The performance of the calibrations

developed in this case were similar to those from Delwiche *et al.* (1998) and again, their application for screening exercises such as found in breeding programmes was proposed.

3.3 NIR spectroscopy

The use of near infrared (NIR) spectroscopy for the determination of cereal composition was pioneered by Karl Norris at the USDA laboratories in Beltsville, USA during the 1960s (Panford, 1987). This early work used transmittance spectroscopy to determine the moisture content of intact wheat kernels and led to the development of the first commercially available NIR instruments. By the early 1970s, instruments based on reflectance principles became available and the applications still used today became established. One of the most important bodies of work in this period was led by Phil Williams at the Canadian Grain Commission (Williams, 1975) which led to the adoption of the technique as an official method for the determination of cereal protein content.

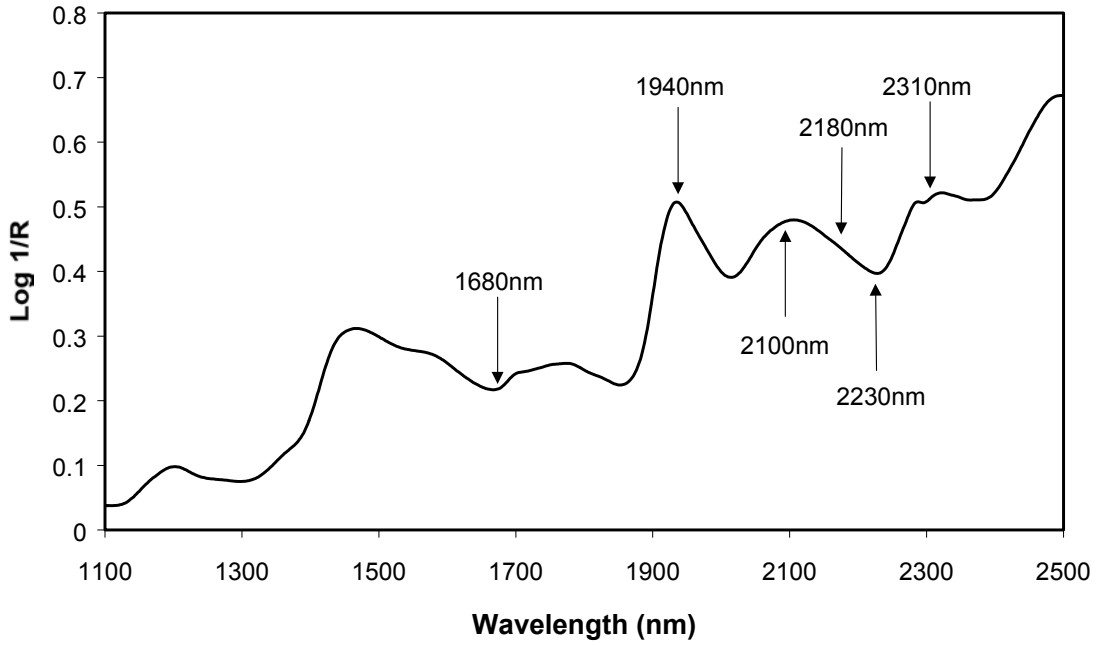
The method was first assessed in the UK at FMBRA in 1975 using UK-grown wheat from the 1974 harvest to produce calibrations for ground wheat protein content (Hart, 1976). There then followed a number of years during which workers at Chorleywood (most prominently Tom Fearn and Brian Osborne) developed calibrations for protein content, moisture content and hardness of wheat (Osborne *et al.*, 1982a; Osborne, 1983, Osborne and Fearn, 1983) and for protein content, moisture content, particle size, colour and starch damage of flour (Osborne *et al.*, 1982b).

3.3.1 Principles of NIR spectroscopy

NIR spectroscopy describes the interaction of electromagnetic energy in the region 800-2500nm with matter. In general terms, NIR instruments contain a source of energy covering the region of interest, a means of irradiating the sample with specific wavelengths or wavelength regions of this light and a detector. The response from the sample at the wavelengths selected is compared with that for a reference material (such as ceramic or gold) and a 'spectrum' for the sample is obtained. Samples having different characteristics produce different spectra and so a means of assessing those characteristics is obtained. Two broad types of instrument exist: those which give information at regular intervals across the entire wavelength range and those which give information about specific, non-continuous regions of the spectrum. Although a variety of methods exist to collect spectral data of the former type, in the work described only one has been used, a scanning monochromator. Instruments of this type use a grating (a highly polished mirror on which are etched a large number of parallel grooves) which disperses the incident energy with respect to wavelength. Historically such instruments have been expensive and so lower cost alternatives have been produced for larger markets. The majority of NIR users in the UK milling industry use instruments which do not produce continuous spectra but rather give information about particular spectral areas by means of interference filters. These so-called filter instruments have proven to be robust and to give acceptable accuracy. An example

spectrum for ground wheat is shown in Figure 1. Although the stated wavelength of typical interference filters are shown, it should be recognised that the information collected for each filter (and indeed each wavelength on the scanning instrument) emanates not just from the centre wavelength but also from wavelengths surrounding it. The term bandwidth is used as a measure of the breadth of the wavelength range of the energy measured for each nominal centre wavelength.

Figure 1. An NIR spectrum for ground wheat collected using a Foss NIRSystems 6500 scanning monochromator with typical interference filter positions indicated



The interaction of NIR energy with matter occurs in two ways. The first is the interaction of specific wavelengths with particular bond types. The main bonds which absorb energy from the NIR region are O-H, N-H and C-H, giving the spectroscopist a means of determining moisture, protein and lipid content respectively. Energy corresponding to particular frequencies (and thus wavelengths) causes these molecules to vibrate and thus absorb incident energy. The fundamental vibrational frequencies for these molecules all fall in the mid infrared region (2500-25000nm) but the overtone and combination absorbance bands arising from these fundamental absorbancies may be found in the near infrared region. It is these bands that may be used to assess sample composition.

In addition to the interaction between electromagnetic radiation and molecules, the interaction of the radiation with the physical structure of a sample also influences the spectrum obtained. In particular, when testing ground wheat or flour, the size of the particles is an important factor to take into consideration. This

is as a result of the greater proportion of particle-air interfaces for smaller particles than large ones. As the radiant energy passes across this interface it may be transmitted directly, refracted or reflected. Where the energy is refracted or reflected, a phenomenon known as scatter occurs. This affects the NIR spectrum collected from a sample because one which is highly scattering, having many small particles, causes the effective pathlength across which the radiation travels to be decreased, i.e. the proportion of the energy absorbed as a result of interaction with molecules is low. NIR instruments which are typically used with fine, particulate materials such as ground wheat respond to the amount of the incident light reflected back from the sample. To allow the spectra obtained to be understood in terms of absorbance regions, the log reciprocal of the proportion of the light reflected is taken, i.e. log 1/R. Samples containing smaller particles scatter more light and produce higher reflectance, which in turn gives lower log 1/R values. The converse is true for materials containing larger particles. In wheat, hard grains tend to produce larger particles when ground and thus generate higher log 1/R values than those which are soft.

3.3.2 Development of NIR calibrations

For virtually all the common applications of NIR in cereal science, it is used as a secondary method, i.e. one which is calibrated against a primary or reference method. Therefore, before an NIR instrument can be used routinely, calibrations for the appropriate measurement methods for the constituents of interest must be developed. These calibrations are simply mathematical equations which relate the responses of the NIR instrument to sample composition. For the vast majority of applications, including those described in this report, this is based on the simple linear regression equation:

$$\hat{Y} = a + bx$$

Where $\hat{Y}$ is the predicted value

a is the intercept constant

b is the multiplicative constant

x is the response used for the prediction

However, although the response at particular wavelengths may be associated with specific constituents such as protein, the physical characteristics of the sample mean that a simple linear regression equation developed for the appropriate wavelengths will not produce the results required. This is because the response at this wavelength will vary not only with sample composition but also with sample particle size and packing density. Typically, therefore, a number of wavelengths are required to produce calibrations which work for a wide range of samples. This requires the development of a multiple linear regression equation:

$$\hat{Y} = a + b_1x_1 + b_2x_2 + \ldots b_nx_n$$

Where $\hat{Y}$ is the predicted value

a is the intercept constant

b_1 to b_n are the multiplicative constants

x_1 to x_n are the responses used for the prediction

It is possible, however, to develop calibrations containing many more constants. While this generally improves calibration accuracy for the calibration set, it can result in the production of calibrations which are unstable and lack general applicability to a range of sample types. For full spectrum instruments, the choice of appropriate sub-sets of wavelengths to produce stable calibrations becomes difficult, due to both the large number of combinations which may be chosen and high correlations between individual wavelengths. To overcome these problems, calibration routines based on the entire spectrum have been developed by adapting other established multivariate statistical techniques. The most well known of these is principal components analysis (PCA) which is used both for calibration purposes and as a basis for discriminating samples on the basis of different spectral characteristics. The basic principle of PCA is that of data reduction. Typically, the spectral differences between different samples result in common differences throughout the spectrum (such as those associated with particle size) and specific differences in particular regions. Therefore a wavelength by wavelength assessment of the spectra is wasteful. By using PCA, the large number of original variables (such as wavelengths) can be reduced to a smaller number of variables called principal components. Each of these describes one direction of variation within the spectra. For ground wheat applications, many hundreds of wavelengths may typically be reduced to 10-20 principal components which still explain >99% of all the variation in the original dataset. In addition to reducing the number of variables needed, PCA also has another advantage. This is related to the way in which principal components are derived. The first principal component is chosen to represent the greatest amount of variation in the dataset, with each subsequent principal component explaining progressively less of the variation. However, the principal components are also chosen such that they are orthogonal or uncorrelated. This means that they can be re-combined in a linear way (such as in a regression analysis) without producing the instability which would result if multiple linear regression were used for a large number of correlated wavelengths. For spectral data for a number of samples, PCA results in the production of principal component loadings which describe the spectral variation for each component and principal component scores which rank each of the samples in the direction of variation represented by each principal component. By linearly combining the products of each of the scores and principal component loadings, the original spectrum for each sample can be reconstructed. In addition to allowing the user to view how different samples relate spectrally to one another, the principal components may also be included in a multiple regression analysis to allow equations to be generated from which the sample composition can be calculated. This process is called principal component regression (PCR). Although this technique works well and can allow the underlying spectral factors associated with the constituent data of interest to be assessed, the technique is not greatly used currently. This is due to the

emergence and adoption of a related regression method which is part of the partial least squares (PLS) suite of analyses. Partial least squares regression (PLSR) is similar in some ways to PCR in that the spectral data are reduced to a smaller number of orthogonal variables. However, rather than using variables derived on the basis of spectral variation only, which, it is recognised, may not represent useful variation in the context of the reference data, PLSR also uses the reference data. The spectral information is reduced in an analogous way to PCA but rather than taking no account of whether they are correlated with the reference data (as in PCA), PLSR ensures that the factors (rather than principal components) derived are correlated with the reference data. Generally this results in regression solutions which use fewer terms (i.e. factors or principal components) than those found using PCR, although the absolute performance of each method is usually comparable. Nevertheless, PLSR does appear to be more widely-used than PCR.

3.3.3 Assessment of NIR calibration performance

When developing calibrations, there are two criteria which should be fulfilled: they should be robust and they should be accurate. A robust calibration is one which is applicable to a wide range of sample types and which is not affected greatly by errors due to instrument noise and different operators. Generally speaking, the more terms included in an equation and the bigger the multiplicative constants used, the less robust the calibration will be (although the accuracy for the calibration sample may be very good). To try to ensure that calibrations are robust, it is usual to include a wide range of samples which vary in respect of all their characteristics, not just those for which calibrations are being derived.

Although a robust calibration is clearly important for long term use, the accuracy of the results obtained is also a key determinant in the calibration used. Ideally the secondary method should give the same results for common samples as the reference method. In reality, however, two factors ensure that this is the never the case. The first of these is the fact that any measurement technique can only ever give an estimate of the true value. This estimate, therefore, includes some error such that when the measurement is repeated, a slightly different value is likely to be recorded. For any joint determination on a common sample, the error for both the reference and NIR methods will ensure that exactly the same result is rarely given. The second error source is the calibration itself as a result of the fact that the spectroscopic and reference techniques typically measure slightly different manifestations of the same parameter.

There are a number of statistical conventions used for describing both the agreement between two datasets and the errors associated with predicting one from the other. Somewhat confusingly, these have a number of different names, depending on the source. There are, however, a number of common concepts which may be used to assess the agreement between NIR and reference results.

For any regression, the correlation coefficient (r for simple linear regression and R for multiple linear regression) is a useful indicator of the closeness of fit for the two datasets:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Where:

x_i = Value predicted by NIR for sample *i*

y_i = Reference value for sample *i*

$\bar{x}$ = Mean of the predicted values

$\bar{y}$ = Mean of the reference values

n = Number of samples

By squaring these figures, the proportion of the total variation which is described by the model is obtained. For NIR predictions of wheat protein content, this figure is usually in the range 0.95-0.99, indicating that this parameter can be predicted accurately using NIR data. Some caution is required in the interpretation of this figure, however, as the range of the *y* values affects the results such that the wider the range covered, the higher the squared correlation coefficient. Thus an apparent improvement may be obtained by including extreme values, even though the actual performance of the calibration for more typical samples is no better. Another disadvantage is that the value given does not allow the user to easily assess what the typical error associated with the result obtained will be. There are a number of different ways of calculating such a figure. The most basic is to calculate the standard error of the regression (*s*) which describes how the predictions are distributed around the regression line (i.e. the line of best fit):

$$s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - b^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{n-2}}$$

Where:

b = Multiplicative constant from regression

This distribution of error is analogous to that obtained when calculating standard deviations. This means that the results obtained will fall within 1 standard error of the regression line approximately 68% of the time,

within 2 standard errors of the regression line approximately 95% of the time and within 3 standard errors of the regression line approximately 99.7% of the time. This statistic basically describes the random variation around the fitted line. However, there may also be systematic differences between the fitted line and the line of equivalence which will affect the accuracy of the results obtained. The two types of systematic errors which may be encountered are slope deviations from the ideal of 1 (skew) and mean difference variations from the ideal of 0 (bias). To assess the total error around the line of equivalence, i.e. $y = x$, the root mean squared deviation (RMSD[1]) is used:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} d_i^{\,2}}{n}}$$

Where:

$d_i =$ Difference between NIR predicted and reference value for sample i

This value tends to be greater than s as both random and systematic sources of error are included in the calculations.

3.3.4 Calibration population structure

Most typically PCA is employed to allow the structure of a population of calibration samples to be assessed (Osborne *et al*., 1993). By plotting the principal component scores in two or three dimensions, the overall shape of the calibration database and thus how the spectra relate to each other may be appreciated.

When using NIR to analyse unknown samples, one of the key factors in assessing how good a prediction may be expected is how similar the spectrum of the unknown sample is to those in the calibration set. Clearly, if it is known that the spectrum of the unknown sample is very different, then the results of the prediction can be judged with that in mind. The statistic which defines how far an observation is from the centre of the calibration set is defined as a Mahalanobis distance or H statistic (Martens and Naes, 1989). When the statistic is applied to principal component scores, it is defined as the geometric distance of each sample from the mean sample in principal component space. The larger this is, the further the observation is from the calibration centre and the greater leverage or influence that observation would have on the calibration if included. This statistic also provides a means of assessing spectral outliers. Typically samples having an H statistic of >3.0 are poorly represented in the calibration set and in the absence of further, similar samples may produce unstable results. Such samples may be removed from the calibration database

[1] Also known as the Standard Error of Prediction (SEP), the Standard Error of Validation (SEV), the Root Mean Squared Error of Prediction (RMSEP) and the Standard Deviation of the Differences (SDD).

to prevent problems in predictive performance. When unknown samples are then measured, their H statistics may then be used to assess whether they are spectrally similar to the calibration samples and where this is not the case, they may be included in further calibration work to ensure ongoing applicability of the calibrations used.

3.3.5 NIR calibration transfer

One of the ongoing problems for users of NIR instrumentation is that of ensuring that calibrations work on instruments other than the one on which the calibration was developed. It is generally accepted that NIR instruments tend to vary in their responses to common samples or materials, even when the instruments used are of the same type and same manufacturer (Dardenne *et al.*, 1992). This is due to small differences in overall photometric response, wavelength scale or wavelength spacing (Tillmann *et al.*, 2000). In turn these can result from small differences in reference materials, lamps, light dispersion or interference systems and detectors. A significant body of work has, therefore, been concerned with the development of suitable means to take account of these differences and thus to develop more efficient ways of maximising the value of calibration work. There are 3 basic approaches to this problem, each of which has merits as well as in some cases significant disadvantages.

- Instrument specific calibrations.
- Calibration transfer by adjustment of predicted results.
- Calibration transfer by mathematical matching of instrument spectral responses.

It is the last of these which has been used in this study as it allows the transfer of the fundamental spectral information rather than simply the predictive aspects. As a result, calibrations may be developed jointly and may be used as a basis for further calibration development through the addition of samples scanned on any instrument as long as the spectra have been suitably adjusted.

In addition, although the simpler adjustment of predicted results approach is widely used for filter instruments, it is recognised that scanning instruments may benefit from a more sophisticated approach. This is because differences between instruments along the wavelength scale can produce errors when full spectrum calibration methods (such as PLSR) are used. Although a number of methods have been proposed to allow the spectral responses of different instruments to be matched, the common goal is to ensure that, for common samples, spectra are obtained which are sufficiently similar to give equivalent results using common calibrations. In addition, spectral outliers should be detected with equal sensitivity on all matched instruments. Although spectral matching approaches are typically more complex than slope and bias approaches, they become important where instruments are run in a network. When large numbers of instruments are supported, it is important that common calibrations can be used to reduce calibration

development and maintenance work as well as to ensure that the performance of individual instruments may be monitored on an equal basis. The main benefits of matching instrument responses in this way are efficiency of calibration development (one calibration should work with a large number of instruments), equal performance for different instruments when assessing spectral outliers, and a means by which new samples may be added to the calibration when required.

In order to derive the mathematical equations or functions required to allow instrument responses to be aligned, measurements of common materials or samples on both the calibration development instrument (the 'master') and the instruments to which the calibrations are to be transferred (the 'slaves') are required. At least one method exists where inert reference materials are used to match instrument responses. This approach is taken by the instrument manufacturing company Bran+Luebbe who use National Institute of Standards and Technology (NIST) standards to allow data transfer between the InfraProver range of instruments. However, other workers (Fearn *et al.*, 1996) have stressed the importance of using spectral features similar to those of the material of interest for the development of calibration transfer algorithms. Arguably the most widely used spectral calibration transfer algorithms for cereals and feedstuffs is that developed by Shenk *et al.* (1985) based on biological samples sealed in cells to ensure stability over time. The original method used 30 samples covering a range of materials, primarily feedstuffs, which had been finely ground. These samples were carefully scanned on both the 'master' and 'slave' instruments and a quadratic model was calculated to align the wavelength reponse on both instruments. Linear regressions were then performed for each of the aligned wavelengths to eliminate differences in photometric response (Shenk *et al.*, 1992). This approach was originally designed for use with NIRSystems instruments (now part of Foss) although it has been applied to monochromator instruments from other manufacturers. Modified algorithms have also been developed for use with filter instruments. More recently, changes in the construction of NIRSystems instruments has allowed a move to a simpler system where a spectral bias is applied to the 'slave' instrument based on analysis of one sealed cell for each product to be assessed, i.e. one for ground wheat, one for flour, one for ground barley and so on. In each case, the sample used is chosen to be near the centre of the population of samples of the product typically encountered (Shenk and Westerhaus, 1995). Samples can be chosen using the PCA and H distance approach described above. This instrument matching approach (commonly called standardisation) is widely used with NIRSystems instruments and has been applied to a number of network situations, particularly in Belgium where a network based on these principles has been used since 1987 (Dardenne and Biston, 1990; Dardenne *et al.*, 1992).

3.4 Project objectives

The overall aim of the work was to develop new NIR calibrations which could be used by the cereals communities in the UK and France and within this, 3 specific objectives were identified.

- To assess the application of size-exclusion high performance liquid chromatography (SE-HPLC) for quality testing of UK and French-grown wheat samples.

- To develop accurate and robust NIR calibrations for wheat and flour quality indices generally and protein quality characteristics, in particular, as measured using SE-HPLC.

- By undertaking a programme of sample exchange to highlight the performance of UK and French-grown wheats under alternative milling, quality testing and baking regimes.

4 Materials and Methods

4.1 Wheat

Samples of wheat representing 3 harvest years (1999-2001 inclusive) were supplied from three sources. The first of these was the annual Recommended and National List assessment undertaken at CCFRA each year from which a total of 181 samples were used. To ensure that samples representing a broad range of properties were selected and to reduce the effect of cross-correlations between related parameters, samples were selected using a semi-factorial approach for each year. The data already collected as part of the work was used as a basis for selection decisions. For breadmaking wheats, samples were selected on the basis of their results for each of the following: flour protein (14% basis), damaged starch, Farinograph water absorption (14% basis), CBP loaf volume, CBP loaf score, no-time dough loaf volume, gel protein mass and gel protein G'. For wheats for biscuit production, samples were selected according to flour protein (14% basis), damaged starch, Farinograph water absorption (14% basis), Extensograph Resistance and Extensibility. Within each category of the samples tested during the study, one sample was selected to represent each of the highest, lowest and the average value for each of these parameters. All the samples were then combined and duplicates removed to produce a set which was representative of the entire range of the properties assessed. Prior to selection, wheat was stored at 12±5ºC in cloth sacks. Following selection, the wheat samples were placed in self-seal polythene bags and returned to reduced temperature storage until required for milling or further analysis.

In addition to those samples described above, 65 samples representing the same three harvest years were supplied from the annual French wheat quality survey by ARVALIS. The remaining 68 samples were obtained from commercial sources including plant breeding companies and flour millers. All of these were also placed in self-seal polythene bags and stored at 12±5ºC until required for milling or further analysis.

A further set of 60 samples from the 2000 and 2001 UK harvests were supplied to ARVALIS where they were assessed in accordance with standard tests of wheat and flour quality and breadmaking performance.

These samples were used to develop combined calibrations and to allow an assessment of the performance of such transferred material under a different quality testing regime.

4.2 Reagents

All of the reagents used were of at least AnalaR grade with the water used for the production of all the solutions used for SE-HPLC analysis being of HPLC grade.

4.3 Milling and grinding of wheat

Wheat samples were Bühler milled to produce white flour in accordance with a standard CCFRA method. Wheat samples were cleaned using a Carter-Day Dockage tester before conditioning to 16% moisture. Normal practice is to condition to 15% moisture and to use milling conditions to generate lower levels of damaged starch for biscuit making varieties. However, in this case, discrimination of breadmaking quality on the basis of differences in protein quality was the priority and so milling conditions similar to those for hard milling wheats were used to ensure that soft milling varieties were not disadvantaged as a result of producing flours of lower water absorbing capacity. Mill settings were optimised in order to achieve flour yields and starch damage levels as close as possible to current commercial practice. The colour of the straight-run flour (break plus reduction mill streams) was measured and where the Flour Colour Grade figure was:

- >2.0, then no bran finisher flour was added.
- between 1.0 and 2.0, a single passage through the bran finisher was carried out and the low grade flour obtained was added.
- less than 1.0, two passages through the bran finisher were carried out and the accumulated low grade flour obtained was added.

A control wheat sample was used to monitor individual mill performance prior to analysis of these test samples. Flour extraction, starch damage and flour colours were recorded each time the control wheat was milled.

A sub-sample of whole wheat was ground using a Perten KT 3100 Falling Number mill which was equipped with a mill feeder (Perten 3170) to produce a wholemeal for NIR analysis. This sample was intended to be representative of the material used for routine analysis of wheat as performed by the majority of UK flour millers.

4.4 Wheat and flour quality assessment at CCFRA

With the exception of protein and moisture content, all quality assessment was carried out as described by Salmon (2001). All Flour Testing Working Group (FTWG) methods are as detailed in CCFRA (2002). A summary of each method used is given below. Except where otherwise described, a proportion of each of the samples analysed in each case was assessed in duplicate to improve the estimate of the true value and to allow calculation of the repeatability of the methods used throughout the work.

4.4.1 Protein and moisture contents

Bühler milled flour and KT ground wheat were assessed for protein content by Dumas (N × 5.7) in accordance with CCFRA FTWG 19. Moisture contents of Bühler milled flour, KT ground wheat and coarsely ground wheat (using a Bühler Miag mill) were determined by oven drying at 130ºC in accordance with FTWG 08.

4.4.2 Particle Size Index

An indication of wheat endosperm texture (hardness) was assessed by air-jet sieve analysis of the milled flour, using a 75µm sieve. The percentage of material passing through the sieve was measured. A cut-off point of 50% is normally used: soft wheat varieties tend to have values above 50% whilst hard wheat varieties produce values below 50%.

4.4.3 Flour colour grade

Flour colour grade was measured using the Kent-Jones & Martin Colour Grader, series 4 (FTWG Method 07/4). The reflection of light at 530nm from a flour/water paste contained in a glass cell was measured. This provides a measure of the bran contamination in a white flour sample, which is related to milling quality. High values tend to indicate greater bran contamination and may be expected to have a detrimental effect on breadmaking quality.

4.4.4 Starch damage

Starch damage was determined according to FTWG method 05. The level of starch damage produced on milling is an important milling quality parameter. The miller aims to control starch damage during milling in order to control water absorption and optimise flour quality for breadmaking.

4.4.5 Hagberg Falling Number

Flour Hagberg Falling Number was measured according to FTWG method 06, the weight of flour used being adjusted according to the moisture content of the flour. This test provides an estimate of the *alpha*-amylase

activity of the flour and this estimate is then used to calculate the amount of fungal *alpha*-amylase required in test baking. (For all baking tests employed in this work total *alpha*-amylase was adjusted to 40 Farrand units (equivalent to 0.8 Ceralpha Units (CU/g)) by addition of a fungal *alpha*-amylase supplement).

4.4.6 Farinograph water absorption

The water absorbing capacity of flour samples was measured using the Brabender Farinograph working to the 600 BU line (FTWG Method 04). This test provides a measure of the amount of water required in order to mix a dough to a fixed consistency. High values are desirable for breadmaking wheats and low for biscuit wheats (the level of water addition indicated by this test is used subsequently in both the CBP and no-time dough test baking processes).

4.5 Test baking

4.5.1 Chorleywood Bread Process (CBP) loaf volume and crumb score

A standard laboratory-scale Chorleywood Bread Process was used to produce 800g four-piece white loaves. This procedure is considered to be closer to commercial practice for basic CBP bread in the UK. Each sample was mixed and baked in duplicate. Loaf volume and the crumb cell structure was measured on a single loaf from each mix. Loaf volume was determined by seed displacement. Before use, the instrument was calibrated using a fibreglass loaf of known volume, 2609ml. The weight of seed displaced by the test loaves was measured and converted to loaf volume.

The traditional method of evaluating crumb structure is to utilise the experience of a trained baker. The assessor ignores single faults which may not be representative of the loaf (e.g. caused by moulding or position of the slicing operation) and, disregarding loaf volume, provides a single score which reflects the average quality over the entire slice cross-section.

Loaves were sliced approximately in the centre of the second piece of the four-piece bread format and a crumb structure score (maximum 10) was assigned on the basis of the visual appearance of the cut surface.

- Crumb structure scores below 5 reflect totally unacceptable bread structure, i.e. crumb structure is very uneven and coarse, and cells are thick walled. Bread in this category normally exhibits poor oven lift and lacks volume, i.e. is rather dense.
- Scores of 5 and 6 are given when the baked product is just bread-like. Crumb structure is poor with large areas of coarse and thick walled cells, visible cores or streaks.
- A crumb score of 7 is assigned to a sample that produces a loaf of acceptable appearance, but exhibits some faults in crumb structure terms, i.e. some areas of coarse, thick walled cells or streaks are evident.

- A score of 8 is assigned to a loaf that exhibits good quality characteristics. The crumb is relatively uniform and the majority of cells are thin walled. No major faults such as cores or streaks are evident.
- A score of 9 is given to a loaf with very good crumb structure, i.e. uniform with thin cell walls and no evidence of major faults.
- A score of 10 is only assigned to the perfect CBP loaf, i.e. where texture consists only of small, uniformly distributed cells in the crumb cross-section and these cells are thin walled.

4.5.2 No-time dough loaf volume

All potential breadmaking samples were also evaluated using a standard no-time dough baking procedure. In this case a spiral mixer was used with mixing being carried out for a fixed amount of time (2 minutes slow Spiral setting and 6 minutes fast) to develop the dough. Single mixes only were performed in the spiral mix system. Measurement of loaf volume was performed as for CBP, but internal crumb structure was not evaluated.

4.6 Gel protein quantity (g/5g flour) and quality (G')

Flour (15g) was defatted with 40ml petroleum ether (b.p. 40-60^{o}C) for 1 hour, filtered and dried. Five grams of defatted flour was stirred with 75ml of 1.8 % sodium dodecyl sulphate (SDS) for 10min at 10^{o}C, then centrifuged at 40,000rpm for 35min. The gel protein layer was removed and weighed. The weight of gel protein represents the amount of functional protein present in the flour. It consists, principally, of glutenin material and is genetically controlled. In general, breadmaking wheats have higher levels than feed or biscuit-making varieties. A typical range for breadmaking would be 9-12g/5g of flour (wet-weight basis). The elastic modulus (G') of gel protein was measured using an Ares Scientific Rheometer and can be used to distinguish between varieties in terms of quality for UK breadmaking. Studies have suggested that samples with a G' of less than 15Pa and those greater than 40Pa may not give optimum performance in the CBP. Excessive gluten strength may also be exhibited as poor performance in the no-time dough baking test.

4.7 Flour protein fractionation and analysis by size-exclusion high performance liquid chromatography (SE-HPLC)

Flour samples were assessed following the Profilblé® method developed jointly by ARVALIS (formerly ITCF) and l'Institut National de la Recherche Agronomique (INRA) (Morel *et al.*, 2000). Flour (160mg) was combined with 20ml of 1% SDS in 0.1M phosphate buffer, pH6.9. This flour in buffer suspension was then agitated to allow the soluble gluten proteins to be extracted from the flour. The suspension was then subjected to a mild sonication treatment using a probe sonicator (Misonix Microson XL2000) to solubilise the polymeric protein fraction. The suspension was then centrifuged for 10min at 5000rpm, after which an aliquot was taken from the supernatant and placed in an HPLC vial. Following sealing, vials were then taken

for SE-HPLC analysis. A total of 32 samples including 2 from a control flour supplied by ARVALIS, could be extracted per day. The control flour was used to allow the results for each run to be evaluated in light of expected performance using a standard control chart approach.

All SE-HPLC analysis was performed using a Jasco system. The analytical column used was a TSKgel G 4000SW and this was operated in conjunction with a TSKgel SW guard column. Samples (20µl) were eluted using 0.1% SDS in 0.1M phosphate buffer with a flow rate of 0.7mlmin^{-1} and detection at 214nm. Each chromatogram took 25min to collect. Chromatograms were integrated using a combination of automatic and manual (using a fixed set of 'rules' developed by ARVALIS) methods depending on the shape of the peak associated with each of the fractions. The response for each sample for a given extraction series was normalised using the results for the control flour for that series. All values are quoted as a percentage of the total response of the column, i.e. as a percentage of total protein.

The overall performance of the columns on the day of analysis was determined by comparing the results for a protein standard (cytochrome C) with control values determined when the column was first used. Each combination of guard and analytical columns was characterised on first use by running 5 protein standards; cytochrome C – also used for subsequent column assessment, carbonic anhydrase, bovine serum albumin, yeast alcohol dehydrogenase and bovine thyroglobulin. The retention time for each was compared with their log molecular weight to assess the linearity of the column's response. A set of flour samples of known protein contents was also used at this stage to develop a calibration curve for the column to allow an assessment of extraction efficiency to be undertaken subsequently. The performance of all the columns used was assessed for each extraction series as described above and the columns replaced when the performance deteriorated below the level recommended in the method developed by ARVALIS.

In addition to the daily and column-specific checks described above, the performance of the method at CCFRA was checked against that at ARVALIS by using a set of samples, analysed at both sites, which represented typical UK and French wheat types.

4.8 Wheat and flour quality assessment at ARVALIS

4.8.1 Milling and grinding of wheat

Wheat was milled in accordance with the BIPEA BY-102-D/9302 procedure using a QUADRUMAT senior mill to produce white flour. Following cleaning using a Rohr mini petkus machine, samples were conditioned to 16% moisture content and were allowed to equilibrate for 24 hours before milling. The extraction rate obtained was between 68 and 72% with flours having an ash content between 0.5 and 0.6 %. The bran produced was passed through a bran finisher twice and the flour collected was added to the flour obtained directly from the mill.

4.8.2 Test baking

A standard laboratotory-scale test was used to assess the baking quality of the samples. This test was developed by the CNERNA (Research national Center on food and nutrition) and produces 5 loaves from 350g of dough. Dough is made using 1500g of flour, yeast (2% on flour weight), salt (2.2% on flour weight) and water (the quantity of water added is judged by the baker during the mixing time). Doughs are then mixed for 20 minutes. After a first fermentation time of about one hour, the dough is divided and moulded into loaves. After a second fermentation (about 1 hour) the loaves are baked for 25 minutes at 250°C.

Every step of the process is scored by the baker to result in a composite total possible score of 300. The behaviour of the dough is scored out of 100 (elasticity, extensibility, consistency during mixing and moulding, behaviour during first and second fermentation and relaxation of the loaves before placing in the oven). The quality of bread is scored out of 200, 100 for the crumb and 100 for the exterior shape of the bread (included the volume). Loaf volume is determined using 2 loaves and the mean result reported. An acceptable volume is between 1500 and 1700ml with values below 1500ml indicating poor performance and those above 1700ml indicating very good performance. The volume is scored out of 30 and the other characteristics of bread are scored out of 70 (colour of the bread, cross-sectional aspect, thickness and appearance). The crumb must be irregular with little and large cells while its texture should be soft and elastic but not sticky. The flavour of crumb should be pleasant. The score for a superior bread making flour should be greater than 215 and 225, for example, would be seen as a very good score. Between 180 and 215, the sample is from a standard breadmaking wheat while below 180, the sample is of poor quality.

4.8.3 Profilblé® results and french bread making performance

To obtain good French bread, flours must produce doughs having a good balance between elasticity and extensibility. Elasticity is required as the bread is produced without a pan and it should retain its shape as it bakes. Extensibility is required to allow the round dough piece to be moulded into a baguette shape which is ~60cm in length. Therefore, the quantity of the protein is not the most important factor compared with the balance between the quantity of gliadins and glutenins. Generally, the ralationship between breadmaking score and protein content is not linear and the optimum balance is dependent on the wheat variety used. With the SE-HPLC method, a ratio of F3+F4/F1 giving a value of about 3 seems to be indicative of good baking potential. The ratio F1/F2 should be in the range 0.5-0.7 with values below this indicating excessive extensibility and values greater indicating excessive elasticity.

4.9 NIR analysis

4.9.1 Instrumentation

All calibration, validation, standardisation and check samples were scanned using a Foss NIRSystems 6500 scanning monochromator spectrometer over the wavelength range 400-2498nm at 2nm intervals. Samples

were presented to the instrument using a sample transport mechanism. Whole wheat samples were presented using the standard rectangular cell supplied by the instrument manufacturer for the analysis of heterogeneous materials. Both ground wheat and flour were placed in a circular cup for scanning. Standardisation and check cells were presented using a proprietary holder for these smaller cells. For each complete spectrum, 32 scans were collected of the ceramic reference material followed by 32 scans of the sample. Sample scans were collected using ISI software. The instrument gain and the linearity of wavelength response were assessed and corrected as necessary once each week. On each day of analysis, the instrument noise and bandwidth characteristics were logged as well as the laboratory temperature and relative humidity.

4.9.2 Sample scanning

Five spectra per sample were collected for whole wheat with 3 being taken for both ground wheat and flour. Whole wheat spectra were collected on day 1 for each set to be analysed (20 per day) before being ground on the following day for both NIR scanning and reference (moisture and protein contents) analysis. As each sample was ground, it was placed in a self-seal polythene bag and left to cool for approximately 1 hour. The sample was then thoroughly mixed and half was transferred to another self-seal polythene bag for subsequent NIR scanning. The remaining sample was placed in another self-seal polythene bag which was retained for subsequent reference testing. Flour samples were scanned by NIR on day 1 of each analysis with the reference analyses being performed on the next day. Previous experience has indicated that strict adherence to this sort of schedule is beneficial, particularly when developing calibrations for parameters sensitive to sample moisture content. However, due to the way in which the samples were collected and analysed as part of the annual quality study prior to the NIR scanning taking place, only the protein and moisture determinations could be coordinated in this way.

4.9.3 NIR instrument standardisation

To allow the development of joint calibrations, the response of the instrument at CCFRA was mathematically aligned (standardised) to that at ARVALIS. For ground wheat and flour analysis, a set of sealed standardisation samples was used. The method was based on that of Shenk and Westerhaus (1995). Sealed cells were analysed using the holder or mechanism of presenting such cells to the instrument as used by the instrument operator. The samples were scanned at CCFRA prior to transportation to ARVALIS for subsequent scanning. The samples were run in a fixed order on each instrument, allowing a minute between scans for the cell to cool. The standardisation samples were always analysed in duplicate and in the same orientation as presented to the instrument. In each case, the cells were rotated by 90º for the second scan to ensure that sample heterogeneity in the cell was taken into account. A total of 33 samples were used covering a range of different types of cereal-derived material to ensure that the range of response obtained on each instrument was sufficient. Of these samples, 23 were used to develop the standardisation equation with the remaining 10 being used to test its performance.

For whole wheat, a selection of 10 samples from the UK database covering a broad range of spectral response was taken. Sub-samples were thoroughly blended prior to division into two, equal portions. One was scanned at CCFRA while the other was transported to ARVALIS where it was also scanned. Five spectra were recorded for each sample on each instrument.

Standardisation equations were developed using the routines included in the WINISI software used for all data analysis.

4.9.4 Data treatment

All reference analysis data were analysed using Microsoft Excel software. All calibration and standardisation data manipulation was carried out using WINISI unless statistical testing of significance for regression analyses was required in which case Microsoft Excel was used with equations entered by the author. All NIR scans were checked against replicates for agreement with the RMSD, bias and RMSD(C) (the RMSD corrected for bias) for the log 1/R values being recorded. In addition, principal components analysis was used to assess how samples were grouped as a function of their spectra. Where inconsistencies were observed, outlying scans were investigated and where warranted, removed. All the valid scans were averaged to produce the spectra used for further data treatment.

Averaged NIR and reference data were saved to one file on which all the calibration or validation work was performed. For full spectrum data, a number of commonly used derivative combinations (1,4,4,1; 2,8,6,1; 3,10,10,1 and 4,10,10,1[2]) recommended by the producers of ISI and WINISI software were assessed for each constituent of interest using both the NIR region only (1100-2498nm) and the whole spectrum (400-2498nm) with wavelength gaps of 2 and 8nm. The effect of Standard Normal Variate & Detrending scatter correction (Barnes *et al.*, 1989, 1993) was assessed for each combination. Calibrations were developed using Modified Partial Least Squares Regression (MPLSR) which is an adaptation of classical PLSR included in ISI software packages. During calibration, outliers were selected and removed in accordance with the software default setting (standardised residual of >2.5). Where a separate validation set was used, samples were selected from the overall file by taking every fourth spectrum and associated reference data and saving to a separate file. The remaining scans were used to develop calibrations which were tested against this set. The scan order was random by constituent and so no systematic errors were introduced as a result. Validation samples having standardised residuals of >3 were removed when calculating and plotting the results. Calibration performance and validation was assessed using the standard statistical approaches outlined above. In

[2] Where the first number indicates the derivative used (1st or 2nd), the second indicates the derivative gap used, the third the primary smoothing gap used and the fourth the secondary smoothing gap (not used in any case and so set to 1).

addition, the ratio of performance to deviation (RPD) was calculated in some cases to facilitate comparison of the performance of calibrations for parameters having very different ranges of values (Williams, 1987).

Calibrations for use with filter instruments were developed by reducing each full spectrum file to the following wavelengths: 540, 1445, 1680, 1940, 2050, 2100, 2139, 2180, 2230 and 2310nm which are typical of those used in the majority of filter instruments for analysis of cereals and their products.

The variance represented within each calibration developed was partitioned by developing separate calibrations based on 1 NIR scan and 1 reference determination. The variance associated with the true model uncertainty was determined using the equation:

$$TTE^2 = \sqrt{(SECV^2 - NIR_r^2 - Ref_r^2)}$$

Where TTE is true test error

SECV is the standard error of cross validation for the calibration developed

NIR_r is the root mean squared difference between NIR results from the same sample obtained using the calibration

Ref_r is the root mean squared difference between reference results from the same sample for the parameter of interest

Partial correlations corrected for protein content were calculated as described by Fearn (1999). Outliers for all the parameters of interest were removed prior to calculation.

4.9.5 Statistical analysis

In addition to the specific software packages already described, general statistical analysis of the reference data, including assessment of cross-correlations and stepwise multiple linear regression analysis, was performed using MINITAB.

5 Results and Discussion

5.1 Wheat and flour properties

One of the most important aspects of NIR calibration processes is the collection and accurate analysis of a range of samples, the properties of which cover those of samples likely to met in the future. This ensures that the calibrations developed will be relevant to the material to be analysed and will be robust, i.e. will give accurate results for a range of different samples. The accuracy and precision of the data presented for calibration, however, also has a major impact on the subsequent performance of the calibration. To ensure

that the range of material collected during this project gave the best possibility for the development of useful calibrations, samples were collected from 3 harvest years to include a range of agronomic and climatic conditions. Samples were then selected from the pool available in each year to represent the maximum, minimum and mean values for each sub-set evaluated as described above. This technique was used to reduce the impact of any correlations, to ensure that a wide range of material was assessed by NIR. In addition, the inclusion of samples grown in France as well as a number of commercial samples in the calibration sets served to further broaden their application. The properties of the samples and an indication of the precision of the reference determinations are given in Table 2.

Table 2. Wheat and flour reference analysis

Material	Parameter	Mean	Minimum	Maximum	RMSD	SD_r
Whole wheat	Protein as is N × 5.7 (%)	10.63	7.95	13.31	0.08	0.05
	Moisture (%)	14.28	11.88	16.2	0.04	0.03
Ground wheat	Protein as is N × 5.7 (%)	10.85	8.1	13.7	0.06	0.05
	Moisture (%)	12.51	10.7	15.5	0.05	0.04
Flour	Protein as is N × 5.7 (%)	9.72	6.8	12.5	0.04	0.03
	Moisture (%)	14.24	12.6	15.6	0.06	0.04
	Farinograph water absorption (%)[a]	56.31	44.8	68.9	0.33	0.24
	Particle Size Index (%)	46.48	26.1	83.9	0.40	0.28
	Hagberg Falling Number (s)	323.4	131	462	13.0	9.2
	Damaged Starch (Farrand Units)	25.0	2	77	1.1	0.8
	Flour Colour Grade (FCG units)	-0.83	-4.7	4.8	0.10	0.07
	Gel protein mass (g)	8.41	1.1	13.8	0.41	0.29
	Gel protein G' (Pa)	33.94	1.7	138.5	3.78	2.67
	CBP loaf volume (ml)	3310.3	2553	3761	55.2	39.0
	CBP crumb score	6.1	3	8	0.6	0.4
	No-time dough loaf volume (ml)	1499.1	1149	1836	-[b]	-[b]
	F1 (%)	13.87	8.6	17.5	0.29	0.21
	F2 (%)	23.69	17.8	26.7	0.48	0.34
	F3 (%)	8.22	6.3	10.8	0.26	0.18
	F4 (%)	38.31	32.1	49.1	0.26	0.19
	F5 (%)	15.89	11.2	218	0.38	0.27
	F1/F2	0.586	0.39	0.77	0.016	0.011
	(F3+F4)/F1	3.432	2.28	6.14	0.098	0.069

[a]50g bowl Farinograph used.
[b]No replication performed for this parameter.

In each case, these results represent approximately the entire range which would be typical for wheat grown in the UK and covers a proportion of that grown in France also. Where a significant extension of any of the ranges was felt to be necessary then it is likely that third country wheats (such as those from North America or Australia) would be required.

The repeatability of the methods used in each case is estimated using SD_r and where values for this statistic have previously been calculated within ring tests (CCFRA, 2002), a comparison of these data with those already known may be carried out. This demonstrates that the precision of the values obtained in this study was similar to that obtained previously, albeit being slightly larger in some cases. This is thought to be a reflection of the protocol for testing, which resulted in duplicates for some of the tests of flour properties being carried out on different days and thus being subject to higher errors. Nevertheless, it was felt that the reference data were sound and represented a good basis for calibration development.

The team at ARVALIS has previously carried out extensive studies on the use of the Profilblé® method and, from this, has adopted limits of precision to assess the quality of the data generated. The work reported here meets these specifications, demonstrating that the underlying reference determinations for all the parameters for which calibrations have been developed are reliable.

5.2 Flour SE-HPLC analysis

Chromatograms had a general form of that illustrated in Figure 2. Peaks indicate time periods across which material of a particular size eluted from the column. Due to the way in which the columns used function, physically larger material eluted earlier than smaller material. Detection was carried out at 214nm, a wavelength at which protein is a strong absorber, and so the area under the chromatogram is a measure of the total protein content of the extract. Assignments of the main features of the chromatogram are indicated on Figure 2 and are derived from development work using other corroborative techniques such as SDS gel electrophoresis (Christine Bar-L'Helgouac'h, personal communication). The region thought to represent HMW glutenins is essentially the void volume of the column, i.e. that material which is excluded throughout the column. The remaining fractions interact with the column material to a greater (larger polymers) or lesser (smaller monomers) degree. The overall shape of the chromatograms obtained were similar to those previously reported (Morel *et al.*, 2000).

An example of the information which may be derived using the technique is given in Figure 3 where flour from 3 UK-grown wheat samples was analysed. The wheats (Soissons, Malacca and mixed Group 4) represent 'extra strong bread', 'strong bread' and 'weak feed' gluten properties respectively. The chromatograms in this case have been normalised for protein content, allowing differences in shape to be observed. The non-breadmaking class clearly shows a significantly lower response for F1 and F2, indicating

the relative lack of large, polymeric proteins. This results in the larger (F3+F4)/F1 ratio results. Although both breadmaking varieties had fairly similar results for (F3+F4)/F1, the difference in their gluten properties is apparent from the difference in F1/F2 ratio observed. That for Soissons was greater than that for Malacca, representing a shift within the glutenins toward relatively more of the higher molecular weight fraction.

Figure 2. Typical SE-HPLC chromatogram for wheat flour protein extracts

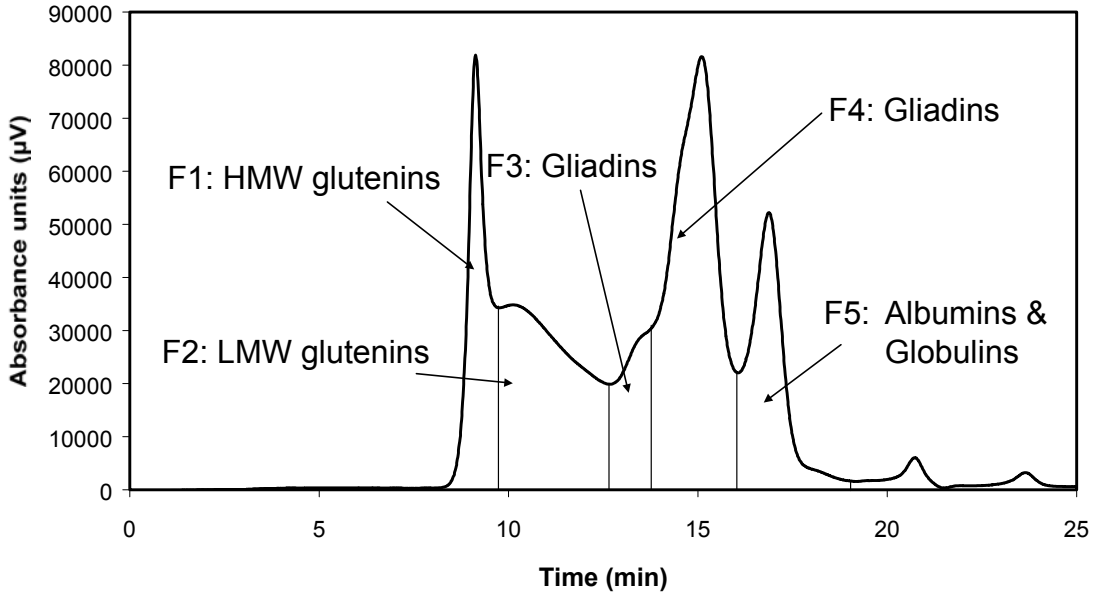
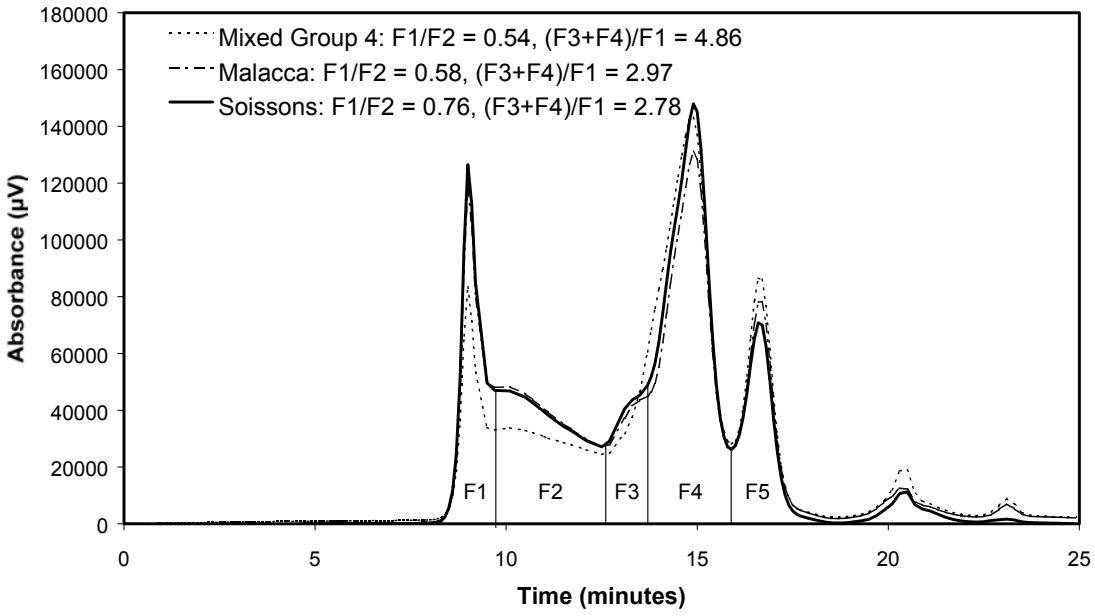


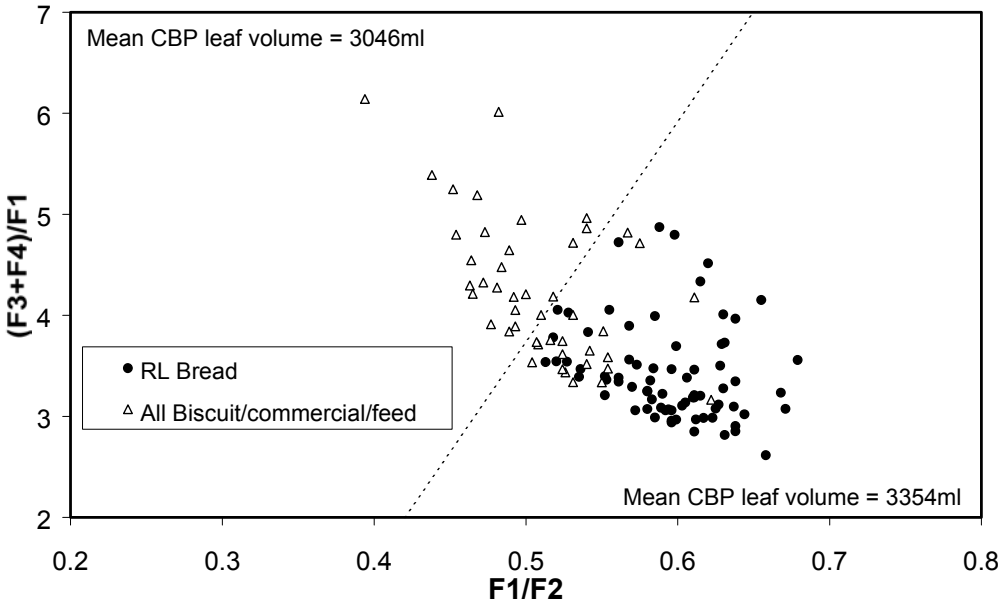
Figure 3. Classification of typical UK-grown varieties using SE-HPLC results (area under chromatogram normalised for protein content)



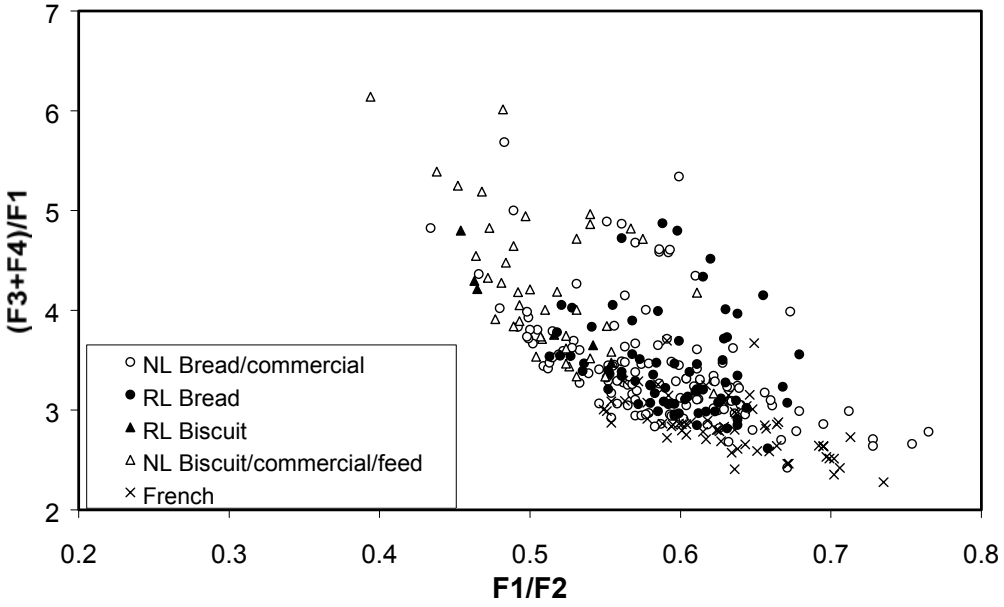
On the basis of this limited comparison, it seems that the (F3+F4)/F1 ratio classifies flours broadly into breadmaking and non-breadmaking classes, while the F1/F2 ratio may be used to discriminate breadmaking wheat into 'extra strong' from 'standard' categories. To understand whether this is consistent for a range of different sample types, the entire dataset as well as the main categories represented within it were grouped using these two ratios (Figure 4).

Figure 4. Classification of UK and French-grown wheats using SE-HPLC results

a. Recommended List bread varieties and all biscuit/unknown commercial and feed varieties



b. All varieties



The first observation arising from this work is the underlying relationship between F1/F2 and (F3+F4)/F1 for the dataset collected. This is not surprising as F1 is a common variable in both ratios and it would be expected that varieties tending to have greater proportions of higher molecular weight material would tend to have higher F1/F2 values and correspondingly lower (F3+F4)/F1 results. To assess how broad categories could be discriminated, the results obtained for all the samples collected, either grown and supplied as biscuit or feed varieties or which were supplied commercially as non-breadmaking, were compared with those from the breadmaking varieties from the Recommended List used in the project (Figure 4a). This shows that the breadmaking varieties generally tend to have higher F1/F2 and lower (F3+F4)/F1 values as would be expected given the fact that they form more elastic doughs and, therefore, that their proteins will tend to produce more extensive cross-linking and may have more material in the higher molecular weight fractions (Southan and MacRitchie, 1999). The samples do demonstrate a range of values, however, with different properties being evident from different relative proportions of higher molecular weight glutenins and gliadin material. There was also some overlap between the groups, with a number of the lower grade wheats giving similar results to those obtained for breadmaking varieties. This would be expected as protein characteristics and baking performance obtained will be based on a continuum even though several sub-classes within this will be evident. To illustrate this, a line separating all the breadmaking varieties from the majority of the samples from the other classes has been drawn, and the performance in CBP test baking on either side of the line has been indicated. Although this is entirely arbitrary, being for illustrative purposes only, and does not reflect the SE-HPLC results alone (protein content will also have a significant effect, for example), the separation is interesting and shows the fundamental nature of the results obtained by the Profilblé® SE-HPLC method and their relationship with flour protein functionality.

When the remaining samples analysed were included (Figure 4b), the range of samples assessed is broadened and extended as the values for F1/F2 increase. This latter effect is primarily due to National List breadmaking samples and a number of the French-grown wheats. A number of the National List breadmaking samples also exhibit lower F1/F2 values than those for the Recommended List samples. These are interesting results and demonstrate the fact that the National List samples will tend to encompass a wider range of properties than those which are finally moved to Recommended List, as would be expected for such a quality selection system. It is also interesting to note the tendency towards values which result in stronger gluten properties for the French-grown samples. Although the majority give similar values to those obtained for UK Recommended List breadmaking varieties, a number give higher F1/F2 and lower (F3+F4)/F1 values. This indicates more elastic gluten properties and may relate to the different requirements for French breadmaking methods. UK breadmaking processes tend to require a different balance of gluten properties in which ‘extra strong’ varieties of this type would be used only as a component of a grist.

5.3 Relationships between measures of flour quality

Given the range of parameters assessed as part of the annual quality assessment and the number of factors which are known to have an impact on breadmaking quality, the correlation between each of the quality measures used was calculated. The results of this exercise are given in Table 3. Each block of two figures comparing the results from two parameters comprises the Pearson correlation between the two parameters and the probability (on a scale of 0 to 1) that the correlation observed may be achieved by chance. As with other measures of statistical significance, values of <0.05 are seen as being significant, those <0.01 are seen as being highly significant and those <0.001 are taken as being very highly significant. Values of >0.05 are classed as not significant. Due to the large number of samples included in the database, even quite small values for the Pearson correlations result in statistically significant relationships, even where they would not be practically useful. To aid in the interpretation of the results, therefore, shading has been added to differentiate those with the highest Pearson correlation values. Perhaps the first and most important parameter which should be compared with the remainder is that of protein content. It has been recognised for many years (Finney and Barmore, 1948) that where other factors such as variety or glutenin subunit composition remain constant then there is a significant positive correlation between loaf volume and protein content. Even where these other factors are not constant, positive correlations may still be found as a certain amount of functional protein is required to produce gluten on hydration and thus to have sufficient material to cover the surface of all the gas bubbles in dough. The gluten films thus formed clearly result in a very high surface area when the number and size of all the gas cells in a typical loaf of bread are considered. It is not surprising, therefore, to see that there is a positive correlation between CBP loaf volume (0.36), CBP crumb score (0.44) and flour protein content. It is more surprising, however, to note the lower correlation observed between flour protein content and loaf volume when produced by a no-time dough method using a spiral mixer. Generally this method of breadmaking is thought to be potentially more discriminating of different flour types than CBP, which tends to realise the potential of flours having lower protein contents. The most strongly correlated parameters with protein content were water absorption, gel protein mass and SE-HPLC fraction F5. The first two of these were positively correlated with protein content; this is not surprising given the well-known relationship between protein content and a flour's capacity for absorbing water (Farrand, 1969) and the accepted dependence of gel protein mass on protein content. In fact, this relationship was one of the reasons that led to the gel protein approach being extended to an assessment of the rheological characteristics of the gel protein fraction (Oliver and Pritchard, 1993). It is interesting to note that the F5 fraction has a negative correlation with protein content. As this fraction is thought to contain the albumin and globulin material, this may be indicative of a positive relationship between protein content and all the functional proteins, i.e. the combined glutenins and gliadins.

Not surprisingly, in addition to protein content, Farinograph water absorption is also positively correlated with damaged starch levels, both of which in turn are negatively correlated with flour particle size index

Table 3. Correlations between flour quality parameters

	Protein	Moisture	WA	PSI	HFN	DS	FCG	Gel mass	G prime	CBP LV	CBP CS	NTD LV	F1	F2	F3	F4	F5	F1/F2
Moisture	0.008 *0.892*																	
WA	0.614 *0.000*	-0.106 *0.061*																
PSI	-0.396 *0.000*	-0.158 *0.005*	-0.593 *0.000*															
HFN	0.316 *0.000*	-0.012 *0.828*	0.226 *0.000*	-0.168 *0.003*														
DS	0.276 *0.000*	0.046 *0.424*	0.720 *0.000*	-0.564 *0.000*	0.166 *0.003*													
FCG	0.227 *0.000*	0.185 *0.001*	0.313 *0.000*	0.078 *0.170*	-0.023 *0.691*	0.258 *0.000*												
Gel mass	0.574 *0.000*	0.124 *0.028*	0.166 *0.003*	-0.154 *0.007*	0.375 *0.000*	-0.021 *0.715*	0.015 *0.794*											
G prime	0.274 *0.000*	-0.167 *0.003*	0.206 *0.000*	-0.325 *0.000*	0.176 *0.002*	0.090 *0.115*	-0.386 *0.000*	0.093 *0.100*										
CBP LV	0.355 *0.000*	-0.018 *0.763*	0.021 *0.720*	-0.213 *0.000*	0.134 *0.020*	-0.033 *0.576*	-0.154 *0.008*	0.385 *0.000*	0.079 *0.176*									
CBP CS	0.435 *0.000*	0.088 *0.131*	0.211 *0.000*	-0.307 *0.000*	0.252 *0.000*	0.090 *0.124*	-0.104 *0.075*	0.427 *0.000*	0.188 *0.001*	0.688 *0.000*								
NTD LV	0.254 *0.000*	-0.063 *0.354*	-0.178 *0.008*	0.141 *0.038*	-0.069 *0.313*	-0.312 *0.000*	-0.137 *0.043*	0.242 *0.000*	-0.239 *0.000*	0.614 *0.000*	0.387 *0.000*							
F1	0.044 *0.445*	0.083 *0.145*	-0.131 *0.022*	-0.188 *0.001*	0.217 *0.000*	-0.083 *0.146*	-0.401 *0.000*	0.434 *0.000*	0.551 *0.000*	0.206 *0.000*	0.163 *0.005*	-0.112 *0.099*						
F2	-0.097 *0.089*	0.105 *0.064*	-0.279 *0.000*	0.086 *0.132*	0.267 *0.000*	-0.275 *0.000*	-0.254 *0.000*	0.591 *0.000*	-0.061 *0.283*	0.097 *0.096*	0.149 *0.010*	0.081 *0.236*	0.527 *0.000*					
F3	0.047 *0.409*	-0.003 *0.955*	0.101 *0.076*	0.125 *0.028*	0.133 *0.019*	-0.018 *0.752*	0.228 *0.000*	0.040 *0.486*	-0.105 *0.065*	-0.274 *0.000*	0.019 *0.750*	-0.177 *0.009*	-0.242 *0.000*	0.112 *0.049*				
F4	0.459 *0.000*	-0.013 *0.819*	0.481 *0.000*	-0.176 *0.002*	-0.139 *0.014*	0.315 *0.000*	0.392 *0.000*	-0.280 *0.000*	-0.101 *0.075*	0.002 *0.967*	0.042 *0.476*	0.089 *0.189*	-0.671 *0.000*	-0.782 *0.000*	-0.031 *0.590*			
F5	-0.764 *0.000*	-0.148 *0.009*	-0.489 *0.000*	0.345 *0.000*	-0.282 *0.000*	-0.197 *0.001*	-0.150 *0.009*	-0.509 *0.000*	-0.242 *0.000*	-0.167 *0.004*	-0.362 *0.000*	-0.052 *0.446*	-0.187 *0.001*	-0.143 *0.011*	-0.251 *0.000*	-0.335 *0.000*		
F1/F2	0.126 *0.028*	0.028 *0.620*	0.042 *0.463*	-0.289 *0.000*	0.075 *0.188*	0.096 *0.092*	-0.299 *0.000*	0.119 *0.036*	0.691 *0.000*	0.177 *0.002*	0.093 *0.109*	-0.182 *0.007*	0.815 *0.000*	-0.060 *0.292*	-0.359 *0.000*	-0.250 *0.000*	-0.133 *0.020*	
(F3+F4)/F1	0.106 *0.065*	-0.073 *0.199*	0.250 *0.000*	0.092 *0.106*	-0.218 *0.000*	0.154 *0.007*	0.432 *0.000*	-0.466 *0.000*	-0.408 *0.000*	-0.198 *0.001*	-0.136 *0.019*	0.065 *0.338*	-0.952 *0.000*	-0.675 *0.000*	0.220 *0.000*	0.823 *0.000*	0.030 *0.600*	-0.664 *0.000*

Pearson correlation 0.5-0.6
Pearson correlation 0.6-0.7
Pearson correlation > 0.7

Normal font indicates Pearson correlation
Italicised font indicates probability

(PSI). Damaged starch level is another component in the Farrand equation for water absorption (Farrand, 1969) with the damaged granules being known to absorb significantly more water by mass than native starch granules. In turn, damaged starch is known to be produced during the reduction process of milling and is a function both of the pressure applied by the rolls during milling and the hardness of the wheat used. As PSI is a measure of hardness and is negatively correlated with it, i.e. harder wheats give lower PSI values, the correlations observed may be easily understood.

One of the interesting aspects of this sort of exercise is the complexity of the relationships, such that any one parameter may be correlated with the others in very different ways. A good example of this is the relationship between gel mass and SE-HPLC fraction F2. Although these parameters are positively correlated, F2 and protein content show no correlation even though gel mass and protein content do. This demonstrates the difficulty of developing one, single test which allows the prediction of a complex parameter (protein quality) which may relate to a number of other parameters in indirect ways. In this instance gel mass and protein content probably do have a causal relationship, perhaps due to a relationship between the proportion of gluten forming proteins and overall protein content. Equally fractions F1 and F2 both show positive correlations with gel mass, as would be expected as they should be measuring the same thing, i.e. the most insoluble proteins. However, neither of these fractions as individual components appear to relate to protein content and this may be simply because they comprise only one aspect of gluten (the glutenins) and take no account of the gliadin fractions.. In addition, F5 is also negatively correlated with gel mass, although in this case the relationship appears more straightforward as the relationship mirrors that for F5 and protein content.

Gel protein G' (elastic modulus) was positively correlated with F1; this is logical given that the use of this rheological assessment of the gel protein fraction was designed to allow differentiation of extra strong varieties (Oliver and Pritchard, 1993). As it is thought that such wheat flours contain relatively higher levels of the higher molecular weight polymers (Southan and MacRitchie, 1999), an increased proportion of F1, the fraction thought to contain the largest polymers, is consistent. The ratio F1/F2 is also positively correlated with G', again probably because of the rheology of the gel protein fraction being a function of the molecular weight distribution of the insoluble gluten proteins.

Not surprisingly, both baking procedures give loaf volume results which are highly correlated and the fact that CBP loaf volume and crumb are also highly correlated is indicative of the way in which common properties (such as good protein characteristics) combine to produce an overall baking quality.

As the SE-HPLC ratios were calculated from the individual values, it is not surprising that there were many inter-correlations between them. It was surprising, however, to see the positive correlation between F2 and F4 and it is difficult to understand why this may be so, as the low molecular weight glutenins and gliadins

would be expected to vary independently of one another. The positive correlation between F1 and F2 is also of interest and may indicate common modes of expression for both glutenin fractions. It would be of interest to use the SE-HPLC technique for the assessment of different sulphur fertilisation regimes as it is thought that the low molecular weight glutenins are more sulphur deficient and thus respond more positively to increased sulphur availability than the high molecular weight glutenins (Southan and MacRitchie, 1999).

Overall, therefore, many of the underlying relationships are as would be expected, although it is useful to be able to quantify their level of importance. It is also encouraging to note that many of the parameters related to functionality clearly do not simply rely on protein content alone. This is partly a function of the complexity of the relationships observed but is also a result of the semi-factorial sample selection procedure which ensured that samples representing high, medium and low values for each parameter, independent of the others, were selected. These results give confidence for NIR calibration development, therefore, as it would seem unlikely that spuriously high performance will be observed as a result of secondary correlations with protein content, a parameter which may be modelled with relative ease using NIR approaches (Delwiche *et al.*, 1998).

An extension of assessing the correlations between individual flour quality indices is the development of predictive models for test baking performance from these parameters. Such an approach has previously been applied to European wheats (Andersson *et al.*, 1994) with limited success. This is probably a result of the degree of complexity of the relationships between the various parameters as discussed above and due to the many elements of flour quality which are known to affect test baking performance as well as the, as yet unquantified, characteristics which clearly also play a significant role. For the flour data collected in this study, stepwise multiple linear regression has been applied to develop a model to enable CBP loaf volume to be predicted using a selection of the other parameters measured. Due to the high inter-correlations and because such a model ideally would be used to allow test baking to be avoided, CBP crumb scores and no-time dough loaf volumes were not included in the model development process. Stepwise multiple linear regression is a technique which aims to find the set of variables which may be combined in a model to predict a parameter of interest. Each variable is included only where its inclusion makes a statistically significant contribution to the overall model. Variables are sequentially added or removed until the error of the model is as low as can be achieved. Using the Stepwise Regression function in Minitab, 9 iterations were performed to arrive at the model with the lowest error. A total of 9 terms were included and this resulted in a model which had a standard error of 161ml for the calibration set and a squared correlation coefficient of 0.39. The model derived had the form:

CBP loaf volume = 4512 + (19.3 × Gel protein mass) - (68.3 × %F3) – (42.3 × Flour Colour Grade) + (87.5 × Protein content) – (2.17 × Gel protein G') – (16.4 × Farinograph water absorption) – (4.65 × PSI) – (51.1 × Moisture content) + (409 × F1/F2)

Those parameters or terms in the model with positive constants will tend to be positively correlated with CBP loaf volume, with the opposite being true for those with negative constants. This appears to give a logical equation when the results from the individual correlations and previous experience are taken into account. Not surprisingly, terms relating to protein content and the amount of gel protein present were positive, although it was surprising that the gel protein G' term was negative especially given that F1/F2 had a positive constant and F3 a negative one. The negative term for Flour Colour Grade indicates that flours with lower bran contents tend to perform better, while that for PSI indicates that harder wheats also produce better performance in bread test baking. This is not surprising given the almost total segregation of hard milling varieties for bread production and soft milling wheats for producing flours for biscuit and cake production. The negative constant for Farinograph water absorption demonstrates again the complexity of the relationships between flour quality parameters as it correlates positively with grain hardness, which in turn has a positive constant. This shows that while grain hardness and loaf volume may be positively correlated as a result of the varietal selection for breadmaking and while harder wheats will tend to produce flours having higher water absorbing capacity, high levels of water absorption in themselves do not result in greater loaf volume (although increases in dough and thus bread yield for a given quantity of flour would be expected). The inclusion of flour moisture content in the model was also surprising and appears to indicate that flours with lower moisture contents tend to contribute positively to increased loaf volume.

The performance of the model thus derived indicates how difficult it is to develop a single unifying method of predicting baking performance. In addition, the level of performance which one could expect with such a model, particularly when attempting to predict unknown samples would be so low that very little confidence could be placed in the results. Of course, if such an exercise resulted in reliable models then such an approach would undoubtedly have already been adopted for the annual quality assessment. The results serve as a useful baseline, however, against which to compare the performance of quality-related NIR calibrations. Where their performance exceeds that found in this case, it may be concluded that there are regions of the spectrum which contribute positively to the measurement of such characteristics and that the calibrations do not simply rely on secondary correlations with other parameters which may be relatively straightforward to assess and which together may be used to effect a reasonable estimate of protein quality and ultimately flour baking potential.

5.4 Repeatability of NIR assessment

When developing NIR calibrations, it is important that at all stages where the variability of the response measured may be controlled, due care is taken to ensure that the data collected are sound. To ensure that representative spectra were collected and that a good estimate of the underlying error associated with the results obtained was determined, spectra of flour and KT-ground wheat were recorded in triplicate, while

those for whole wheat were taken in quintuplicate. Each scan was compared with the others to allow an assessment of the underlying agreement between the scans acquired to be performed (Table 4).

Table 4. Agreement between NIR scans (1100-2498nm) collected for inclusion in the calibration database (n=313 for flour and 314 for both wheat types)

Material	Comparison	RMSD (μAu[a])	Bias (μAu[a])	RMSD(C) (μAu[a])
Flour	Scan 1 v Scan 2	2396	168	964
	Scan 1 v Scan 3	2246	107	948
	Scan 1 v Scan 3	2091	-61	900
	Average[b]	**2248**	**71**	**938**
Ground wheat	Scan 1 v Scan 2	4344	-68	1778
	Scan 1 v Scan 3	4628	-48	2041
	Scan 1 v Scan 3	4729	19	1945
	Average[b]	**4570**	**-32**	**1924**
Whole wheat	Scan 1 v Scan 2	12107	-893	4756
	Scan 1 v Scan 3	11617	-834	4579
	Scan 1 v Scan 4	11243	-348	4611
	Scan 1 v Scan 5	10015	-55	4326
	Scan 2 v Scan 3	12518	59	4708
	Scan 2 v Scan 4	12308	546	4702
	Scan 2 v Scan 5	11856	838	4711
	Scan 3 v Scan 4	10579	487	4197
	Scan 3 v Scan 5	10502	779	4279
	Scan 4 v Scan 5	10337	292	4248
	Average[b]	**11340**	**87**	**4517**

[a]Micro-absorbance units.

[b]Absolute average (sign disregarded) for bias; root, mean, square of individual values for RMSD and RMSD(C).

It should be noted that these results are in micro log 1/R units and so an overall bias of 71 equates to a difference of 0.000071 log 1/R units. This is insignificant when compared with the log 1/R values recorded in the spectra, which can reach ~0.8 for flour and ground wheat and higher again for whole wheat. When the bias for individual samples has been removed, the RMSD(C) figure gives an indication of the total random error remaining. This is a good indicator of poor packing, where an erroneous sample has been scanned or where a change in sample moisture status has occurred. The results obtained in each case were as would be

expected for an experienced operator. The results according to heterogeneity of the material assessed, with the results for flour being the lowest (i.e. the spectra collected were the most similar) and those for whole wheat being the highest. The values obtained in this last case illustrate the importance of collecting a number of scans for such material to obtain the best estimate of the true value, as was carried out in this study.

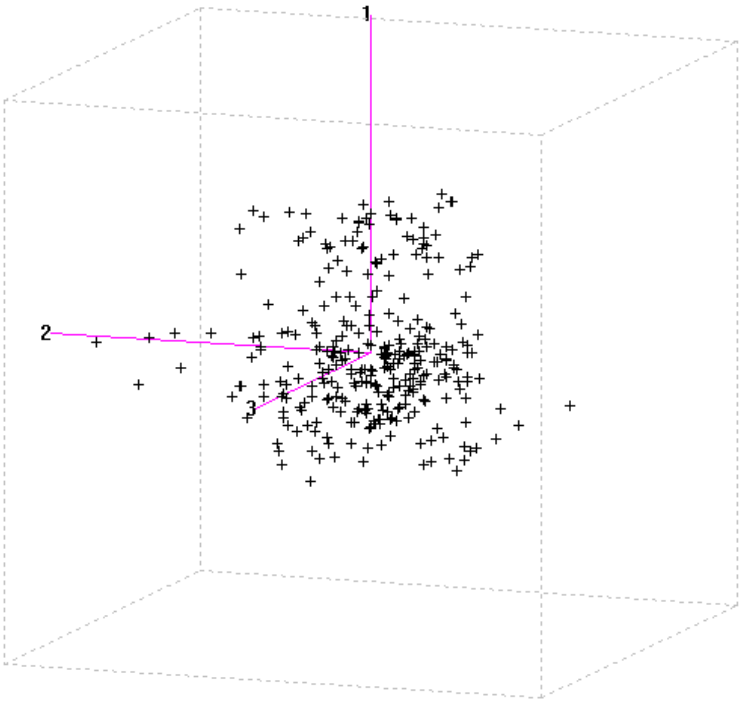
5.5 NIR calibration database development

One important element of ensuring that the samples included in an NIR calibration set are representative is to understand the variability of the samples both in terms of the reference results to be calibrated against but also the variability of the spectra for these samples. A large number of wheat and flour characteristics affect the shape of the NIR spectrum collected, including grain colour (for whole grain), flour particle size, moisture content, protein content, lipid content and so on. Where spectra representing samples having spectral characteristics likely to be found in routine use are not included in a calibration set, then any calibrations developed from these sets will tend to lack robustness and may give poor performance for such samples. To assess the structure of the calibration sets for each of the materials assessed, principal components analysis was performed. The results of this exercise are given for the first 3 principal components (which describe the majority of the spectral variation) for each material (Figure 5). The graphs show that the samples selected gave reasonably symmetrical populations in each case. In each situation, samples may be characterised by the position in the principal component space which they occupy. The distance from the mean position (the centre of the population) is defined as Global H where the H statistic is derived from the Mahanalobis distance. The H statistic is standardised to be analogous to standard deviations from a normal distribution, i.e. values of >2 indicate borderline outliers (typically found 1 time in 20 by chance) while values of >3 indicate significant outliers (typically found 3 times in a 1000 by chance). Very few of the samples included in each database fell into the outlier category for either statistic and given the good coverage of the principal component space for each, it is felt that a good representation of wheat and flour samples has been collected.

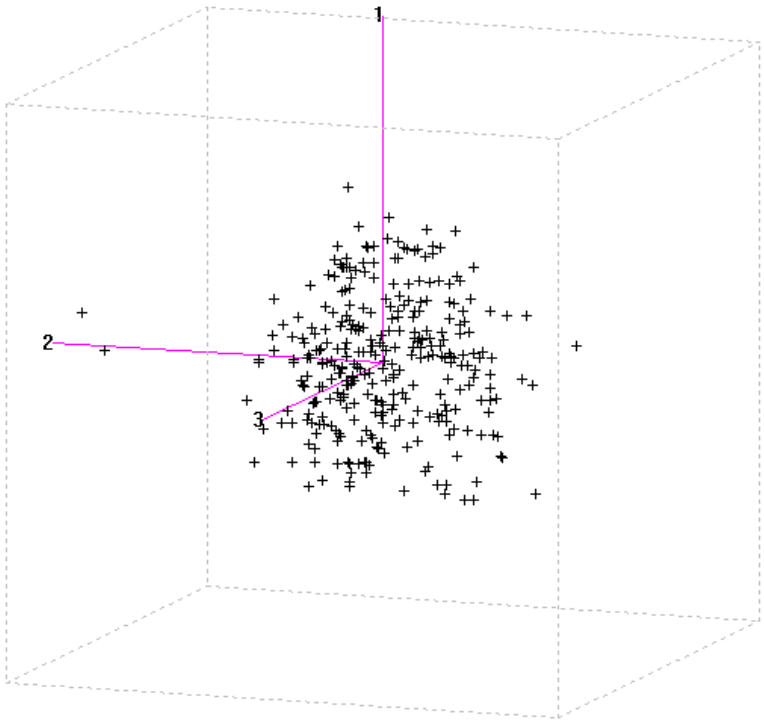
The only samples which were found to be consistently outlying were two samples of wheat having a different bran coloration which were apparent for spectra from both ground wheat and whole wheat and which may be seen in Figures 5b and 5c. The spectra from these samples were very similar in the NIR region, as would be expected as their compositions and particle size when ground or milled were similar to the other wheats in the calibration database. However, their spectra in the visible region were very different (Figure 6) as a result of their different bran pigmentation.

Figure 5. Principal components analysis of NIR calibration databases

a. Flour



b. Ground wheat



c. Whole wheat

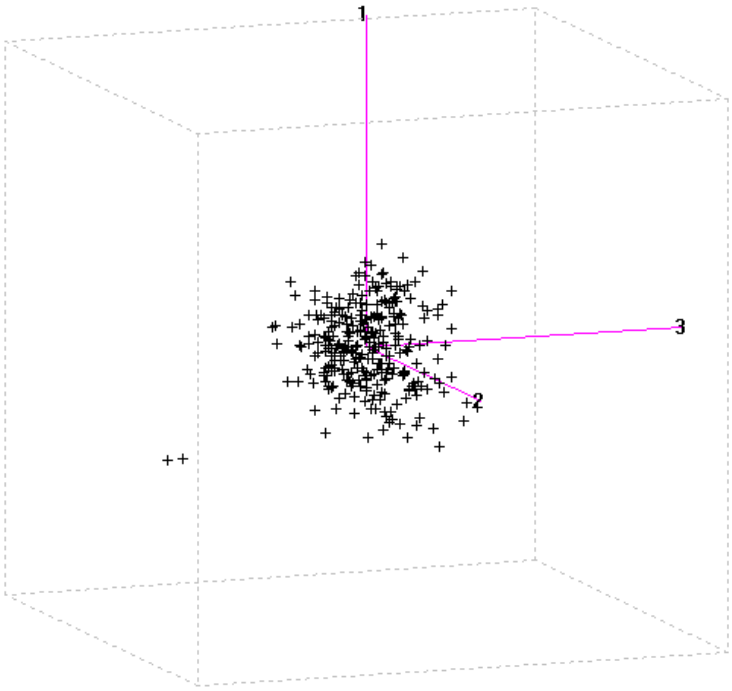
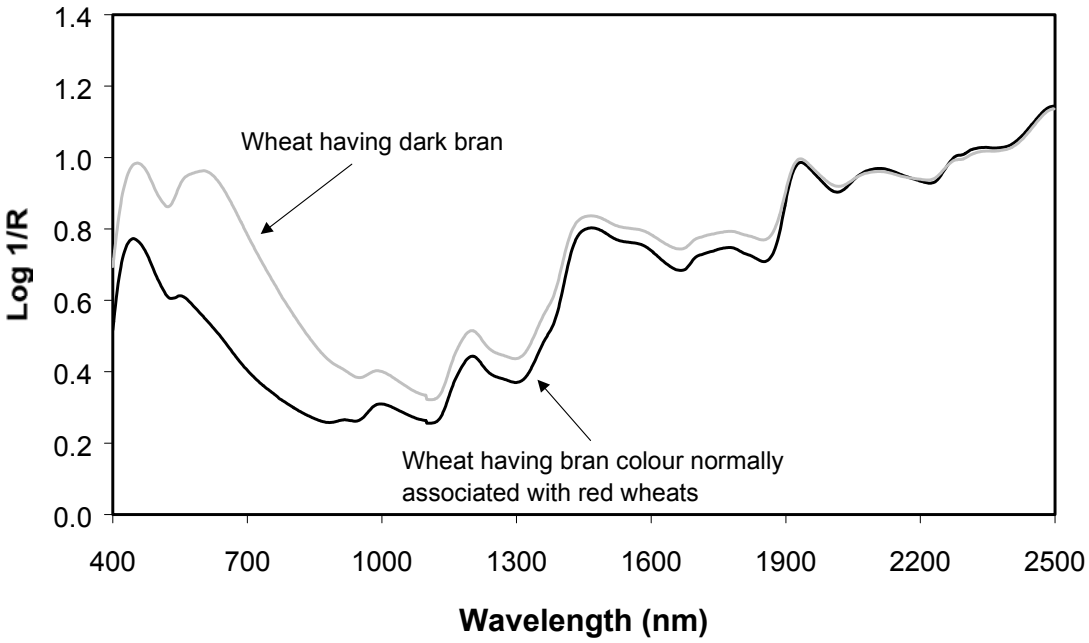


Figure 6. Example spectra showing effect of bran coloration on NIR spectra of whole wheat



5.6 Selection of optimum NIR data pre-treatment routines

Due to the range of sample properties which contribute to the NIR spectrum, data pre-treatments are commonly used, particularly with full spectrum instruments of the type used here, to ensure that the most accurate and robust models result. These treatments generally fall into two categories: a derivative and

smoothing transformation, and the use of mathematical approaches for the reduction of particle size and pathlength effects. In addition to these, the spectral region used for calibration development as well as the datapoint gap have also been varied in this case. Due to the large number of combinations of different approaches, selection of the optimum routine is time-consuming, particularly with larger datasets. As a result, all the data treatments for assessment were used to develop calibrations using the samples collected for the 1999 harvest crop only. The SECV results for each of the combinations for each of the materials are given in Tables 5-7.

The results indicate that there were global differences for the different materials, which in most cases were consistent. In each case, protein content may be used to allow an overall comparison to be drawn, particularly as it is a parameter for which NIR calibrations may be generated relatively easily. It can be seen that there was a ranking of whole wheat which gave larger SECVs than ground wheat which, in turn, gave larger values than flour. As one of the aims of NIR calibration is to reduce this statistic to a minimum, it is clear that the results for flour were better than those for wheat. There was also some indication that, for a number of the parameters assessed, the range of error values obtained ranked calibrations in a similar way; i.e. not only were the whole wheat calibrations the poorest but obtaining good performance was more difficult generally. This trend is reasonably consistent for all the parameters assessed, although some exceptions were observed. The results for protein content were generally very good, being significantly lower in all cases than the limits adopted in the UK (ICC, 1986). The reason for the differences between the materials is most probably a function of the more homogeneous nature of flour than ground wheat and whole wheat respectively. This was also noted for the assessment of repeatability of the NIR spectral data as discussed above. This was also the reason for collecting 5 scans for each sample of whole wheat, whereas only 3 were used for both ground wheat and flour. There was a very consistent pattern of response to different data pre-treatments for protein content: the results obtained when a scatter correction treatment was applied (SNV + DT) were superior to those without it. Again, this seems to indicate that decreasing the information related to particle size or pathlength in the spectra appears to benefit protein calibration development.

The results for moisture content were interesting, as in this case, there was less of an effect of scatter correction pre-treatment and smaller differences between the different materials. In fact, here the whole wheat calibrations were slightly superior to those obtained for ground wheat. It is thought that this is because the spectral response to moisture is relatively much larger than that for protein and so is less affected by other spectral features. The effect of sample handling is also more significant for moisture than for other parameters. This is because even small variations (~0.1%) in moisture status due to interaction with the surrounding environment will affect NIR calibration development, whereas the effect such small variations in moisture content would have on protein content or water absorption, for example, is much more limited.

Table 5. Effect of data pre-treatments on NIR calibration performance for flour (standard errors of cross-validation)

Math treatment	Gap	Region	SNV+DT	Protein	Moisture	WA	PSI	HFN	DS	FCG	Gel mass	G prime	CBP LV	CBP CS	NTD LV	F1	F2	F3	F4	F5	F1/F2	(F3+F4)/F1
1,4,4,1	2	Both	No	0.12	0.10	1.40	2.18	39.1	2.94	0.40	1.23	17.15	164.2	0.7	111.3	1.15	1.22	0.55	1.81	0.61	0.043	0.439
1,4,4,1	2	Both	Yes	0.05	0.09	1.45	2.02	36.0	3.45	0.34	1.23	16.54	172.0	0.7	118.5	1.11	1.18	0.48	1.84	0.58	0.041	0.424
1,4,4,1	2	NIR only	No	0.11	0.09	1.24	2.01	36.8	2.87	0.59	1.15	17.44	155.5	0.8	112.3	1.13	1.30	0.57	1.78	0.60	0.046	0.397
1,4,4,1	2	NIR only	Yes	0.05	0.09	1.42	2.13	36.3	3.56	0.53	1.15	19.08	155.7	0.7	104.1	1.18	1.34	0.51	1.69	0.63	0.048	0.398
1,4,4,1	8	Both	No	0.12	0.09	1.44	2.09	41.9	3.16	0.40	1.27	17.10	165.9	0.7	110.8	1.20	1.22	0.55	1.76	0.65	0.043	0.440
1,4,4,1	8	Both	Yes	0.05	0.09	1.46	2.28	35.9	3.57	0.34	1.23	15.75	173.0	0.7	117.1	1.15	1.19	0.48	1.83	0.59	0.041	0.442
1,4,4,1	8	NIR only	No	0.13	0.09	1.40	2.01	37.1	3.03	0.53	1.16	17.33	156.3	0.8	103.1	1.15	1.32	0.50	1.80	0.65	0.046	0.391
1,4,4,1	8	NIR only	Yes	0.05	0.08	1.37	2.18	35.6	3.40	0.52	1.13	18.96	155.1	0.7	102.9	1.16	1.33	0.50	1.70	0.63	0.048	0.402
2,8,6,1	2	Both	No	0.12	0.10	1.52	2.43	42.5	3.54	0.40	1.09	17.01	169.7	0.7	127.4	1.19	1.54	0.56	2.04	0.70	0.041	0.474
2,8,6,1	2	Both	Yes	0.06	0.09	1.48	2.30	42.8	3.73	0.37	1.05	16.91	155.1	0.7	119.1	0.97	1.43	0.54	2.11	0.63	0.041	0.419
2,8,6,1	2	NIR only	No	0.12	0.09	1.43	2.20	41.6	3.38	0.61	1.18	19.33	148.2	0.8	99.4	1.19	1.32	0.49	1.60	0.61	0.046	0.358
2,8,6,1	2	NIR only	Yes	0.06	0.09	1.49	2.23	40.6	3.75	0.58	1.13	19.88	143.6	0.7	112.7	1.08	1.22	0.47	1.58	0.61	0.046	0.358
2,8,6,1	8	Both	No	0.12	0.09	1.52	2.43	39.1	3.06	0.43	1.12	17.57	164.5	0.7	115.7	1.08	1.16	0.54	1.73	0.62	0.040	0.417
2,8,6,1	8	Both	Yes	0.06	0.09	1.50	2.64	37.6	3.87	0.33	1.09	16.90	157.4	0.7	108.1	1.02	1.10	0.53	1.60	0.57	0.039	0.365
2,8,6,1	8	NIR only	No	0.12	0.09	1.43	2.29	39.5	3.26	0.60	1.17	19.28	149.0	0.8	111.0	1.18	1.25	0.49	1.63	0.62	0.049	0.425
2,8,6,1	8	NIR only	Yes	0.06	0.08	1.48	2.22	40.6	3.75	0.58	1.17	20.39	143.8	0.7	112.5	1.09	1.23	0.47	1.61	0.62	0.045	0.383
3,10,10,1	2	Both	No	0.12	0.09	1.52	2.22	42.9	3.74	0.50	1.15	16.43	169.0	0.8	112.6	1.09	1.31	0.59	2.02	0.66	0.043	0.458
3,10,10,1	2	Both	Yes	0.06	0.09	1.46	2.34	43.5	3.61	0.41	1.10	15.16	157.3	0.7	116.5	1.05	1.36	0.52	1.65	0.62	0.042	0.427
3,10,10,1	2	NIR only	No	0.11	0.10	1.39	2.13	40.3	3.31	0.62	1.22	19.47	157.2	0.7	102.0	1.09	1.23	0.51	1.54	0.61	0.049	0.387
3,10,10,1	2	NIR only	Yes	0.05	0.09	1.50	2.30	44.9	3.82	0.65	1.20	20.41	150.3	0.8	115.0	1.13	1.10	0.47	1.47	0.60	0.047	0.366
3,10,10,1	8	Both	No	0.12	0.10	1.50	2.65	36.0	3.28	0.47	1.14	16.14	163.4	0.8	99.5	1.06	1.14	0.56	1.64	0.58	0.037	0.395
3,10,10,1	8	Both	Yes	0.06	0.09	1.46	2.63	44.7	3.95	0.40	1.10	15.29	159.5	0.7	109.1	1.02	1.18	0.54	1.44	0.59	0.039	0.373
3,10,10,1	8	NIR only	No	0.12	0.09	1.45	2.37	40.6	3.33	0.65	1.22	19.49	156.8	0.7	101.9	1.09	1.24	0.50	1.55	0.61	0.049	0.391
3,10,10,1	8	NIR only	Yes	0.05	0.09	1.50	2.32	43.8	3.83	0.65	1.21	20.44	150.9	0.8	96.6	1.05	1.10	0.47	1.46	0.60	0.050	0.363
4,10,10,1	2	Both	No	0.12	0.09	1.48	2.18	42.4	3.89	0.45	1.08	18.35	166.5	0.8	123.3	1.15	1.40	0.65	1.98	0.73	0.047	0.447
4,10,10,1	2	Both	Yes	0.06	0.10	1.49	2.09	42.4	4.01	0.41	1.06	17.55	154.6	0.8	119.8	0.97	1.39	0.59	1.71	0.67	0.047	0.431
4,10,10,1	2	NIR only	No	0.12	0.08	1.47	2.27	42.2	3.41	0.66	1.17	19.68	144.5	0.8	107.3	1.04	1.17	0.48	1.50	0.64	0.047	0.392
4,10,10,1	2	NIR only	Yes	0.05	0.09	1.49	2.28	44.7	3.91	0.60	1.23	19.86	139.5	0.7	116.3	1.06	1.20	0.48	1.58	0.61	0.049	0.422
4,10,10,1	8	Both	No	0.09	0.09	1.48	2.22	40.4	3.41	0.45	1.07	17.59	168.9	0.7	117.2	1.13	1.12	0.54	1.46	0.62	0.046	0.405
4,10,10,1	8	Both	Yes	0.06	0.10	1.48	2.31	42.0	3.85	0.36	1.04	19.40	155.7	0.7	116.2	1.16	1.12	0.52	1.45	0.59	0.047	0.445
4,10,10,1	8	NIR only	No	0.12	0.09	1.47	2.26	40.0	3.41	0.62	1.17	19.69	148.3	0.7	106.8	1.06	1.26	0.49	1.52	0.62	0.049	0.405
4,10,10,1	8	NIR only	Yes	0.06	0.09	1.48	2.25	43.9	3.88	0.60	1.20	19.88	138.9	0.7	122.8	1.10	1.21	0.45	1.45	0.59	0.047	0.428
			Minimum	**0.05**	**0.08**	**1.24**	**2.01**	**35.6**	**2.87**	**0.33**	**1.04**	**15.16**	**138.9**	**0.7**	**96.6**	**0.97**	**1.10**	**0.45**	**1.44**	**0.57**	**0.037**	**0.358**
			Maximum	**0.13**	**0.10**	**1.52**	**2.65**	**44.9**	**4.01**	**0.66**	**1.27**	**20.44**	**173.0**	**0.8**	**127.4**	**1.20**	**1.54**	**0.65**	**2.11**	**0.73**	**0.050**	**0.474**

For mathematical treatment the first number indicates the derivative used (1st or 2nd), the second indicates the derivative gap used, the third the primary smoothing gap used and the fourth the secondary smoothing gap (not used in any case and so set to 1). The gap is in nanometres.
SNV+DT is Standard Normal Variate with Detrending Treatment to reduce particle size and pathlength effects.
WA is Farinograph water absorption, PSI is Particle Size Index, HFN is Hagberg Falling Number, DS is Farrand Damaged Starch, FCG is Flour Colour Grade, CBP LV is Chorleywood Bread Process loaf volume, CBP CS is Chorleywood Bread Process crumb score and NTD LV is no-time dough loaf volume.
The lowest errors for the simplest data pre-treatment option are highlighted in grey.

Table 6. Effect of data pre-treatments on NIR calibration performance for ground wheat (standard errors of cross-validation)

Math treatment	Gap	Region	SNV+DT	Protein	Moisture	WA	PSI	HFN	DS	FCG	Gel mass	G prime	CBP LV	CBP CS	NTD LV	F1	F2	F3	F4	F5	F1/F2	(F3+F4)/F1
1,4,4,1	2	Both	No	0.12	0.13	2.00	3.14	39.0	3.3	0.68	1.23	18.97	176.3	0.7	120.7	1.30	1.29	0.61	2.09	0.69	0.049	0.468
1,4,4,1	2	Both	Yes	0.10	0.14	1.76	2.92	38.3	3.5	0.65	1.18	17.72	158.0	0.7	109.6	1.29	1.46	0.62	2.10	0.68	0.048	0.518
1,4,4,1	2	NIR only	No	0.13	0.13	1.98	2.71	44.3	3.8	0.71	1.31	17.93	169.2	0.7	109.7	1.37	1.61	0.59	2.04	0.68	0.053	0.457
1,4,4,1	2	NIR only	Yes	0.08	0.14	1.94	2.76	49.5	3.8	0.70	1.34	19.32	174.7	0.7	91.3	1.45	1.39	0.56	2.06	0.61	0.045	0.448
1,4,4,1	8	Both	No	0.12	0.13	1.78	3.13	39.1	3.3	0.68	1.23	18.98	176.3	0.7	119.1	1.30	1.30	0.60	2.10	0.69	0.050	0.468
1,4,4,1	8	Both	Yes	0.10	0.14	1.75	2.88	38.4	3.5	0.65	1.18	17.71	158.1	0.7	109.5	1.29	1.49	0.62	2.10	0.66	0.048	0.518
1,4,4,1	8	NIR only	No	0.14	0.13	1.99	2.99	44.5	3.9	0.71	1.30	18.87	170.6	0.7	112.2	1.34	1.46	0.60	2.02	0.66	0.051	0.432
1,4,4,1	8	NIR only	Yes	0.08	0.14	1.93	2.78	49.5	3.8	0.71	1.34	19.30	176.6	0.7	92.2	1.36	1.43	0.55	2.09	0.60	0.045	0.447
2,8,6,1	2	Both	No	0.13	0.14	1.89	3.00	38.7	3.6	0.69	1.19	19.77	161.1	0.7	103.3	1.30	1.26	0.62	2.00	0.75	0.052	0.485
2,8,6,1	2	Both	Yes	0.11	0.13	2.01	3.06	38.8	3.7	0.69	1.19	20.05	166.4	0.7	106.2	1.33	1.14	0.54	2.03	0.68	0.050	0.480
2,8,6,1	2	NIR only	No	0.14	0.15	2.06	2.58	43.7	4.0	0.74	1.24	20.96	165.1	0.7	109.2	1.19	1.44	0.60	1.91	0.70	0.053	0.429
2,8,6,1	2	NIR only	Yes	0.10	0.14	1.92	2.77	43.3	4.0	0.73	1.30	19.68	175.4	0.7	93.1	1.41	1.51	0.57	1.97	0.64	0.050	0.444
2,8,6,1	8	Both	No	0.13	0.14	1.90	3.05	36.6	3.6	0.69	1.16	19.52	161.3	0.7	102.9	1.30	1.26	0.62	1.98	0.72	0.051	0.483
2,8,6,1	8	Both	Yes	0.11	0.13	2.01	3.04	38.1	3.5	0.69	1.20	19.71	166.4	0.7	106.3	1.34	1.13	0.55	2.08	0.68	0.050	0.479
2,8,6,1	8	NIR only	No	0.14	0.15	1.94	2.64	43.7	4.0	0.74	1.25	20.97	165.3	0.7	109.1	1.27	1.43	0.60	1.90	0.70	0.053	0.431
2,8,6,1	8	NIR only	Yes	0.11	0.14	1.92	2.76	44.9	3.8	0.74	1.29	19.72	173.7	0.7	95.1	1.29	1.54	0.57	1.92	0.64	0.050	0.445
3,10,10,1	2	Both	No	0.13	0.14	2.02	3.72	46.3	3.7	0.68	1.26	19.86	162.8	0.7	110.8	1.41	1.49	0.61	1.89	0.74	0.052	0.478
3,10,10,1	2	Both	Yes	0.11	0.13	1.99	2.88	41.2	4.0	0.71	1.10	18.22	164.2	0.7	106.0	1.28	1.23	0.59	1.86	0.60	0.052	0.476
3,10,10,1	2	NIR only	No	0.14	0.12	1.99	2.60	46.3	4.1	0.72	1.26	20.66	169.5	0.7	108.6	1.35	1.41	0.62	1.82	0.69	0.053	0.488
3,10,10,1	2	NIR only	Yes	0.10	0.13	2.02	2.72	43.8	4.2	0.72	1.27	19.83	174.5	0.7	94.8	1.45	1.49	0.63	1.84	0.67	0.050	0.465
3,10,10,1	8	Both	No	0.13	0.13	2.02	3.67	47.1	3.7	0.68	1.22	19.83	163.6	0.7	101.6	1.42	1.48	0.61	1.88	0.74	0.052	0.489
3,10,10,1	8	Both	Yes	0.11	0.14	1.93	2.87	43.2	4.0	0.71	1.11	18.16	167.9	0.7	105.8	1.27	1.23	0.59	1.88	0.60	0.052	0.476
3,10,10,1	8	NIR only	No	0.14	0.12	1.99	2.61	45.3	4.0	0.72	1.30	20.73	165.8	0.7	100.0	1.31	1.47	0.63	1.86	0.69	0.053	0.484
3,10,10,1	8	NIR only	Yes	0.10	0.14	1.88	2.55	44.0	4.2	0.70	1.29	19.77	175.1	0.7	95.7	1.45	1.50	0.61	2.03	0.67	0.050	0.464
4,10,10,1	2	Both	No	0.13	0.15	2.00	3.01	45.4	3.6	0.71	1.21	19.43	172.5	0.7	117.2	1.36	1.42	0.61	1.94	0.74	0.053	0.501
4,10,10,1	2	Both	Yes	0.11	0.14	1.95	2.89	41.7	4.0	0.72	1.17	20.14	156.3	0.7	111.1	1.37	1.29	0.60	1.92	0.63	0.053	0.488
4,10,10,1	2	NIR only	No	0.15	0.13	2.22	2.46	45.3	3.9	0.74	1.22	20.98	172.5	0.7	103.8	1.37	1.33	0.64	1.78	0.77	0.053	0.449
4,10,10,1	2	NIR only	Yes	0.10	0.14	2.12	2.79	44.8	4.2	0.71	1.16	19.50	171.2	0.7	96.0	1.39	1.40	0.62	1.71	0.74	0.050	0.436
4,10,10,1	8	Both	No	0.14	0.15	2.01	2.94	47.4	3.7	0.71	1.20	19.42	172.5	0.7	117.8	1.35	1.35	0.61	1.93	0.74	0.053	0.482
4,10,10,1	8	Both	Yes	0.11	0.15	2.01	2.92	40.7	4.0	0.72	1.22	20.35	158.2	0.7	110.2	1.35	1.30	0.61	1.82	0.63	0.053	0.486
4,10,10,1	8	NIR only	No	0.15	0.12	2.11	2.40	47.1	4.0	0.73	1.15	21.00	172.6	0.7	106.7	1.32	1.38	0.64	1.76	0.79	0.053	0.472
4,10,10,1	8	NIR only	Yes	0.10	0.14	2.20	2.81	43.6	4.2	0.71	1.25	19.95	172.1	0.7	97.1	1.37	1.41	0.63	1.66	0.69	0.050	0.435
			Minimum	**0.08**	**0.12**	**1.75**	**2.40**	**36.6**	**3.3**	**0.65**	**1.10**	**17.71**	**156.3**	**0.7**	**91.3**	**1.19**	**1.13**	**0.54**	**1.66**	**0.60**	**0.045**	**0.429**
			Maximum	**0.15**	**0.15**	**2.22**	**3.72**	**49.5**	**4.2**	**0.74**	**1.34**	**21.00**	**176.6**	**0.7**	**120.7**	**1.45**	**1.61**	**0.64**	**2.10**	**0.79**	**0.053**	**0.518**

For mathematical treatment the first number indicates the derivative used (1st or 2nd), the second indicates the derivative gap used, the third the primary smoothing gap used and the fourth the secondary smoothing gap (not used in any case and so set to 1). The gap is in nanometres.
SNV+DT is Standard Normal Variate with Detrending Treatment to reduce particle size and pathlength effects.
WA is Farinograph water absorption, PSI is Particle Size Index, HFN is Hagberg Falling Number, DS is Farrand Damaged Starch, FCG is Flour Colour Grade, CBP LV is Chorleywood Bread Process loaf volume, CBP CS is Chorleywood Bread Process crumb score and NTD LV is no-time dough loaf volume.
The lowest errors for the simplest data pre-treatment option are highlighted in grey.

Table 7. Effect of data pre-treatments on NIR calibration performance for whole wheat (standard errors of cross-validation)

Math treatment	Gap	Region	SNV+DT	Protein	Moisture	WA	PSI	HFN	DS	FCG	Gel mass	G prime	CBP LV	CBP CS	NTD LV	F1	F2	F3	F4	F5	F1/F2	(F3+F4)/F1
1,4,4,1	2	Both	No	0.21	0.14	2.25	6.15	42.5	5.6	0.68	1.21	17.15	165.6	0.8	128.6	1.60	1.56	0.67	2.36	0.81	0.049	0.542
1,4,4,1	2	Both	Yes	0.19	0.12	2.32	6.37	43.4	5.4	0.61	1.28	18.49	166.8	0.7	115.5	1.59	1.41	0.66	2.37	0.94	0.048	0.542
1,4,4,1	2	NIR only	No	0.21	0.13	2.13	6.88	45.7	6.2	0.72	1.35	18.16	166.0	0.8	126.0	1.60	1.63	0.67	2.50	0.82	0.054	0.526
1,4,4,1	2	NIR only	Yes	0.19	0.12	2.19	6.43	44.2	6.1	0.70	1.27	17.03	177.5	0.9	120.5	1.59	1.54	0.67	2.65	0.94	0.046	0.520
1,4,4,1	8	Both	No	0.21	0.14	2.24	6.14	42.5	5.6	0.67	1.20	17.13	165.6	0.8	128.7	1.60	1.42	0.67	2.36	0.80	0.049	0.549
1,4,4,1	8	Both	Yes	0.19	0.12	2.32	6.37	43.7	5.4	0.60	1.31	17.59	166.6	0.7	115.5	1.59	1.49	0.66	2.37	0.91	0.048	0.495
1,4,4,1	8	NIR only	No	0.22	0.13	2.05	6.88	45.8	6.2	0.72	1.35	18.32	166.0	0.8	126.0	1.60	1.63	0.67	2.55	0.82	0.053	0.512
1,4,4,1	8	NIR only	Yes	0.18	0.12	2.20	6.42	45.6	5.8	0.70	1.27	16.88	177.4	0.7	120.6	1.59	1.53	0.67	2.65	0.94	0.050	0.517
2,8,6,1	2	Both	No	0.18	0.13	2.34	6.55	47.1	5.7	0.63	1.17	17.67	165.1	0.8	137.3	1.61	1.33	0.66	2.31	0.78	0.051	0.528
2,8,6,1	2	Both	Yes	0.15	0.12	2.25	6.05	48.3	5.4	0.64	1.23	17.01	170.9	0.8	126.3	1.59	1.65	0.68	2.25	0.75	0.045	0.561
2,8,6,1	2	NIR only	No	0.21	0.13	2.28	6.97	49.7	5.6	0.75	1.27	19.55	166.0	0.8	128.0	1.52	1.74	0.66	2.51	0.80	0.055	0.503
2,8,6,1	2	NIR only	Yes	0.17	0.12	2.28	6.07	46.1	6.1	0.74	1.29	20.19	170.1	0.8	123.7	1.58	1.66	0.68	2.29	0.81	0.045	0.518
2,8,6,1	8	Both	No	0.19	0.13	2.27	6.55	46.6	5.7	0.63	1.18	17.19	165.1	0.8	137.4	1.61	1.33	0.66	2.31	0.73	0.051	0.528
2,8,6,1	8	Both	Yes	0.15	0.12	2.25	6.06	48.3	5.4	0.64	1.23	17.05	170.9	0.8	126.2	1.59	1.65	0.68	2.24	0.74	0.045	0.561
2,8,6,1	8	NIR only	No	0.21	0.13	2.28	6.71	49.7	5.7	0.75	1.28	19.54	166.0	0.8	128.2	1.52	1.74	0.66	2.51	0.80	0.062	0.503
2,8,6,1	8	NIR only	Yes	0.16	0.12	2.28	6.07	45.9	6.1	0.74	1.29	20.19	170.1	0.8	124.0	1.58	1.66	0.67	2.29	0.81	0.045	0.518
3,10,10,1	2	Both	No	0.20	0.13	2.22	6.88	46.1	5.5	0.62	1.21	15.60	165.9	0.8	134.7	1.50	1.38	0.67	2.23	0.82	0.056	0.526
3,10,10,1	2	Both	Yes	0.16	0.11	2.16	6.12	46.1	5.5	0.61	1.18	15.22	169.7	0.8	120.6	1.53	1.45	0.66	2.20	0.85	0.053	0.531
3,10,10,1	2	NIR only	No	0.20	0.13	2.23	7.21	48.0	5.9	0.74	1.29	19.45	169.4	0.8	124.7	1.53	1.78	0.66	2.51	0.74	0.060	0.501
3,10,10,1	2	NIR only	Yes	0.16	0.11	2.21	6.37	50.1	5.9	0.67	1.39	20.23	169.0	0.8	114.3	1.51	1.67	0.68	2.27	0.87	0.048	0.463
3,10,10,1	8	Both	No	0.21	0.13	2.21	6.88	46.2	5.6	0.62	1.22	15.65	165.8	0.8	116.8	1.50	1.38	0.67	2.24	0.79	0.057	0.526
3,10,10,1	8	Both	Yes	0.16	0.11	2.16	6.12	46.1	5.5	0.61	1.25	15.16	169.7	0.8	121.3	1.53	1.45	0.66	2.20	0.86	0.053	0.506
3,10,10,1	8	NIR only	No	0.20	0.13	2.22	7.02	48.1	5.9	0.74	1.29	19.42	164.7	0.8	125.0	1.53	1.78	0.66	2.49	0.76	0.060	0.537
3,10,10,1	8	NIR only	Yes	0.17	0.11	2.28	6.37	48.4	5.9	0.67	1.38	20.22	169.1	0.8	114.8	1.38	1.67	0.68	2.27	0.86	0.048	0.463
4,10,10,1	2	Both	No	0.19	0.14	2.28	6.84	47.1	5.5	0.65	1.15	17.78	166.2	0.7	131.6	1.55	1.47	0.67	2.37	0.76	0.055	0.627
4,10,10,1	2	Both	Yes	0.19	0.12	2.08	6.35	45.1	5.4	0.60	1.42	17.10	168.6	0.8	113.4	1.43	1.47	0.68	2.11	0.88	0.046	0.583
4,10,10,1	2	NIR only	No	0.22	0.14	2.19	6.96	47.3	5.6	0.75	1.23	19.36	171.3	0.8	123.7	1.56	1.59	0.67	2.39	0.77	0.053	0.536
4,10,10,1	2	NIR only	Yes	0.20	0.11	2.24	6.37	49.7	5.9	0.72	1.36	21.02	175.6	0.8	114.3	1.42	1.67	0.67	2.25	0.85	0.045	0.555
4,10,10,1	8	Both	No	0.20	0.15	2.29	6.85	47.2	5.5	0.63	1.15	17.49	166.3	0.7	131.1	1.49	1.48	0.67	2.36	0.76	0.055	0.612
4,10,10,1	8	Both	Yes	0.20	0.12	2.09	6.35	46.6	5.5	0.63	1.15	17.19	168.7	0.8	113.5	1.43	1.46	0.68	2.35	0.85	0.046	0.583
4,10,10,1	8	NIR only	No	0.23	0.12	2.19	6.96	47.3	6.1	0.75	1.26	19.37	171.4	0.8	123.6	1.45	1.59	0.67	2.39	0.76	0.053	0.533
4,10,10,1	8	NIR only	Yes	0.20	0.11	2.24	6.64	49.0	5.9	0.68	1.22	21.04	175.9	0.8	114.3	1.41	1.67	0.67	2.24	0.86	0.045	0.464
			Minimum	**0.15**	**0.11**	**2.05**	**6.05**	**42.5**	**5.4**	**0.60**	**1.15**	**15.16**	**164.7**	**0.7**	**113.4**	**1.38**	**1.33**	**0.66**	**2.11**	**0.73**	**0.045**	**0.463**
			Maximum	**0.23**	**0.15**	**2.34**	**7.21**	**50.1**	**6.2**	**0.75**	**1.42**	**21.04**	**177.5**	**0.9**	**137.4**	**1.61**	**1.78**	**0.68**	**2.65**	**0.94**	**0.062**	**0.627**

For mathematical treatment the first number indicates the derivative used (1st or 2nd), the second indicates the derivative gap used, the third the primary smoothing gap used and the fourth the secondary smoothing gap (not used in any case and so set to 1). The gap is in nanometres.
SNV+DT is Standard Normal Variate with Detrending Treatment to reduce particle size and pathlength effects.
WA is Farinograph water absorption, PSI is Particle Size Index, HFN is Hagberg Falling Number, DS is Farrand Damaged Starch, FCG is Flour Colour Grade, CBP LV is Chorleywood Bread Process loaf volume, CBP CS is Chorleywood Bread Process crumb score and NTD LV is no-time dough loaf volume.
The lowest errors for the simplest data pre-treatment option are highlighted in grey.

In this case, handling whole wheat is less likely to induce such changes as the surface area presented to the atmosphere is less than that when using finely ground wheat.

The results for Farinograph water absorption once again rank in accordance with material heterogeneity, although in this and subsequent cases, it is important to note that the reference values were generated by analysis of the flour, while the NIR spectra were from those of the material of interest. Therefore, only the flour calibrations represented the same material being analysed in both cases. As a result, the performance of the wheat calibrations would always be expected to be inferior due to the intermediate processing, the effect of which on final flour properties will not be evident in the NIR spectra. The results for PSI further indicate the importance of this point. It is known that NIR spectra of ground wheat and flour may be used to determine particle size with considerable accuracy (Hareland, 1994). However, reports of the use of spectra collected from whole wheat for the assessment of wheat endosperm texture are more limited in number and tend to indicate poor potential for the technique with such material (Delwiche, 1993). This is confirmed here; the errors for PSI were much greater for whole wheat than for the ground and milled material. The same trend may be observed for damaged starch and again this is probably due to the relative lack of information in the whole wheat spectra about endosperm texture and, thus, ultimately, the particle size distribution of milled or ground products.

The results for flour Hagberg Falling Number again show ranking of the results from the different materials, but in this case, such results were virtually meaningless as the absolute level of performance was very poor. This is not surprising given that the test responds to and measures the reduction in starch gel viscosity due to the action of cereal *alpha*-amylase. As this is only one protein in a material which contains many different proteins, it is thought unlikely that NIR could be used to effect a direct measurement.

The results for Flour Colour Grade were surprising in that, although the spectra from flour produced the best calibrations as would be expected, those for ground wheat and flour gave very similar performance. As it is not just the colour of the bran (which the spectrum of wheat in the region 400-2500nm would be sensitive to) which will affect this but also the quantity, it is intriguing to note that the relationship is as good as it appears to be. As a means of measuring the likely colour of resultant flour could be of interest in assessing wheat prior to milling, particularly for management of extraction rates, this could be an interesting result for flour millers.

The results for gel protein mass followed the same trend as many of the other parameters, with progressive improvements in performance as the wheat becomes more highly processed. However, it is interesting to note that this was not the case for gel protein G', where the results for ground wheat were markedly poorer than those for either whole wheat or flour. In addition, the results for whole wheat were very similar to those for flour. This would seem to indicate that the spectral information necessary to develop this calibration is

present all through processing (as the underlying property is) but that grinding using a high speed grinder (Perten KT 3100 Falling Number mill) has a detrimental effect on NIR calibration performance. This is interesting as it has been proposed that the use of this type of mill should be avoided when performing sedimentation determinations of gluten quality (whether SDS or Zeleny) (Williams *et al.*, 2000). As the extraction procedure for gel protein is based on similar principles to that of SDS sedimentation, it may be proposed that damage either to gluten properties or to those elements in the NIR spectrum which may be related to them resulted in the drop in performance observed.

Clearly the basis for the calibrations for test baking performance are not the same as for gel protein since the results for ground wheat did not appear to be adversely affected. In fact, the results for loaf volume when using a no-time dough baking approach were better for ground wheat than for the other materials. In addition, this was not confined to one treatment which out-performed the others but rather appeared to be consistent for a number of different data pre-treatment approaches. In each case, however, the results for test baking parameters were quite poor and indicate the problems inherent in developing calibrations for a complex response which depends on a significant amount of additional processing, even in the case of flour.

It is interesting to note that, for the protein quality-related parameters, the higher derivatives tended to produce better results for flour. The use of third and fourth derivatives for analysis of wheat and wheat flour has not been widely reported, although such treatments have been applied successfully to the analysis of feedstuffs. However, there is a precedent for their use for the prediction of complex quality indicators (Wesley *et al.*, 1998a) and in this case, it would seem that their use is justified given the results obtained. It is usual practice, however, to develop NIR calibrations on the basis that the model developed should only be as complex as required to carry out the task required of it. Higher derivatives clearly have a role in aiding the definition of the minor peaks within the spectrum but some care is also required in their use due to the increased spectral noise that also tends to result. It is for this reason that larger smoothing gaps (10 datapoints) have been used to compensate for such problems.

The results for all the SE-HPLC parameters were consistent, in that the closer the NIR spectrum used for calibration was to the material on which the SE-HPLC was performed, then the better was the calibration performance. Overall the performance of these calibrations was encouraging, particularly when compared with the calculated reproducibility data for the SE-HPLC method (F1: 1.2%, F2: 1.9%, F3: 0.7%, F4: 0.9%, F5: 1.0%). A range of different data pre-treatments were found to give good performance and no obvious patterns were apparent in the way that they were for protein content. It might be expected that the results for the F1 and F1/F2 parameters would have been similar to those for gel protein G' given the correlations observed between them as described above. The performance for calibrations derived using ground wheat spectra, however, did not show any evidence of poor performance when compared with those for whole

wheat in the way that was observed for gel protein G', and so it may be concluded that different underlying relationships were responsible for the different parameters.

5.7 Calibration performance for the entire calibration database

Calibrations were developed using all the best data pre-treatments determined above for all the samples for which NIR spectra and reference results had been obtained. In each case, cross-validation was used to allow the optimum number of terms for inclusion to be assessed and to give standard error values which were more meaningful in terms of likely future predictive potential.

In each case, the results when all the samples from all three years were included (Tables 8-10) tended to follow similar trends in terms of material and parameter assessed to those from the first year's harvest (1999) used for determination of the optimum data pre-treatment. Generally the results for flour were better than those for ground wheat and, in turn, those for whole wheat. Although SEC and SECV may not be easily used to directly compare results for parameters having different scales, the squared correlation coefficient may be used to give a 'rule of thumb' assessment of the relative performances for each of the parameters. This figure is derived from the calibration set, however, and so suffers from the same shortcoming as SEC in that it is likely to be a poor measure of future predictive performance. To compensate for this, therefore, another statistic, RPD (ratio of standard error of performance to standard deviation) has been calculated (Williams, 1987). In this case, SECV has been used rather than the standard error of performance and has been compared with the standard deviation of the reference values obtained for each of the parameters assessed (Figure 7). Where values of approximately 1 (i.e. SECV = standard deviation of reference method) are found, then it may be concluded that the calibration has no greater predictive performance than simply choosing a result at random from the range normally expected. Increasing values indicate progressively better performance. The results for calibrations derived using flour (Figure 7a) indicate that, for the majority of the parameters assessed, reasonable performance was obtained. Overall the poorest calibrations were those for Hagberg Falling Number, gel protein G' and test baking parameters. The result for protein content of flour, however, had the highest RPD that the author has seen reported for wheat and flour testing. This indicates the very high quality of this calibration, confirmed by the SECV (Table 8), which was 0.05%, whereas the limit currently adopted in the UK for the determination of protein content in flour is 0.2% (ICC, 1986).

The results obtained for most of the predictors of flour functionality indicated that NIR has clear potential to be used as a screening tool for protein quality determinations. This is less marked for the parameters highlighted above as having low RPDs, but for gel protein gel mass and the HPLC characteristics the results were very promising.

Table 8. Calibration performance for the entire flour calibration set

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content as is (%)	291	9.70	1.20	>0.99	0.05	0.05
Moisture content (%)	295	14.25	0.53	0.97	0.09	0.10
Water absorption (%)	301	56.27	3.64	0.93	1.00	1.14
Particle Size Index (%)	296	46.11	9.13	0.94	2.32	2.57
Hagberg Falling Number (s)	303	326.6	51.6	0.51	36.2	42.0
Damaged starch (FU)	295	24.7	8.8	0.93	2.4	2.7
Flour Colour Grade	301	-0.86	1.48	0.93	0.39	0.41
Gel protein gel mass (g)	303	8.50	1.93	0.79	0.89	0.97
Gel protein G' (Pa)	297	32.37	20.48	0.70	11.31	13.16
CBP loaf volume (ml)	286	3326.0	185.9	0.62	114.6	128.4
CBP crumb score	292	6.1	1.1	0.54	0.7	0.8
No-time dough loaf volume (ml)	212	1500.4	122.2	0.56	81.4	95.8
F1 (%)	301	13.87	1.64	0.79	0.75	0.92
F2 (%)	303	23.68	1.68	0.64	1.00	1.18
F3 (%)	301	8.23	0.74	0.77	0.36	0.42
F4 (%)	303	38.30	3.02	0.83	1.26	1.44
F5 (%)	297	15.88	1.70	0.93	0.46	0.52
F1/F2	303	0.584	0.058	0.64	0.035	0.040
(F3+F4)/F1	302	3.432	0.646	0.82	0.277	0.326

[a]Standard error of calibration
[b]Standard error of cross validation

Table 9. Calibration performance for the entire ground wheat calibration set

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content as is (%)	285	10.89	1.10	>0.99	0.08	0.10
Moisture content (%)	300	12.51	0.71	0.96	0.14	0.15
Water absorption (%)	294	56.33	3.56	0.77	1.71	1.75
Particle Size Index (%)	293	46.14	9.13	0.83	3.71	3.94
Hagberg Falling Number (s)	304	327.0	52.3	0.43	39.6	44.4
Damaged starch (FU)	295	24.8	8.8	0.82	3.7	4.0
Flour Colour Grade	301	-0.86	1.48	0.80	0.67	0.85
Gel protein gel mass (g)	301	8.47	1.95	0.74	0.99	1.12
Gel protein G' (Pa)	296	31.91	19.46	0.52	13.51	14.55
CBP loaf volume (ml)	286	3326.6	186.1	0.40	143.8	156.7
CBP crumb score	294	6.1	1.1	0.21	1.0	1.0
No-time dough loaf volume (ml)	208	1505.4	121.9	0.58	78.9	86.4
F1 (%)	297	13.86	1.63	0.74	0.84	1.02
F2 (%)	299	23.70	1.67	0.65	0.98	1.17
F3 (%)	301	8.24	0.75	0.72	0.40	0.51
F4 (%)	299	38.32	2.99	0.76	1.46	1.69
F5 (%)	296	15.85	1.71	0.91	0.52	0.59
F1/F2	299	0.585	0.058	0.55	0.038	0.044
(F3+F4)/F1	302	3.426	0.650	0.74	0.334	0.398

[a]Standard error of calibration
[b]Standard error of cross validation

Table 10. Calibration performance for the entire whole wheat calibration set

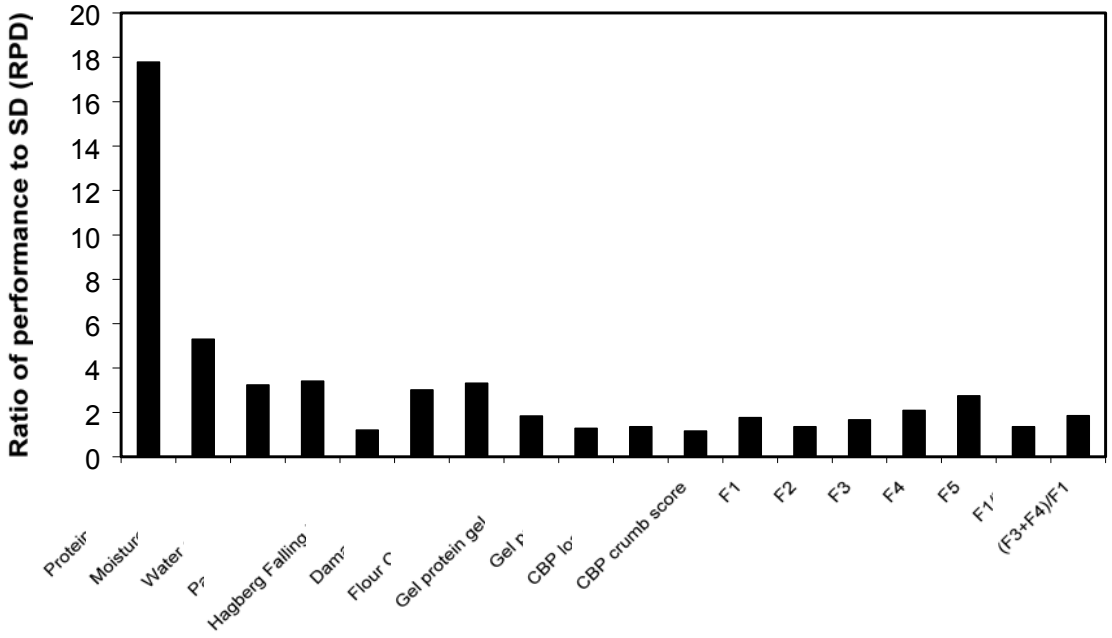
Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content as is (%)	284	10.66	1.06	0.99	0.12	0.15
Moisture content (%)	293	14.30	0.91	0.99	0.08	0.10
Water absorption (%)	299	56.27	3.66	0.68	2.06	2.19
Particle Size Index (%)	295	45.77	8.87	0.50	6.28	7.19
Hagberg Falling Number (s)	300	326.0	51.3	0.26	44.2	46.3
Damaged starch (FU)	296	24.9	9.0	0.67	5.2	5.9
Flour Colour Grade	298	-0.87	1.43	0.74	0.73	0.80
Gel protein gel mass (g)	299	8.42	1.94	0.72	1.03	1.21
Gel protein G' (Pa)	293	31.19	18.33	0.53	12.60	13.91
CBP loaf volume (ml)	282	3332.0	185.2	0.41	142.7	148.5
CBP crumb score	284	6.2	1.0	0.26	0.9	0.9
No-time dough loaf volume (ml)	211	1507.4	123.7	0.37	97.9	103.9
F1 (%)	298	13.87	1.65	0.48	1.19	1.27
F2 (%)	291	23.86	1.46	0.23	1.28	1.32
F3 (%)	299	8.22	0.73	0.42	0.55	0.60
F4 (%)	300	38.24	2.99	0.72	1.57	1.88
F5 (%)	294	15.86	1.70	0.87	0.60	0.84
F1/F2	298	0.584	0.058	0.57	0.038	0.045
(F3+F4)/F1	296	3.386	0.609	0.49	0.433	0.454

[a]Standard error of calibration
[b]Standard error of cross validation

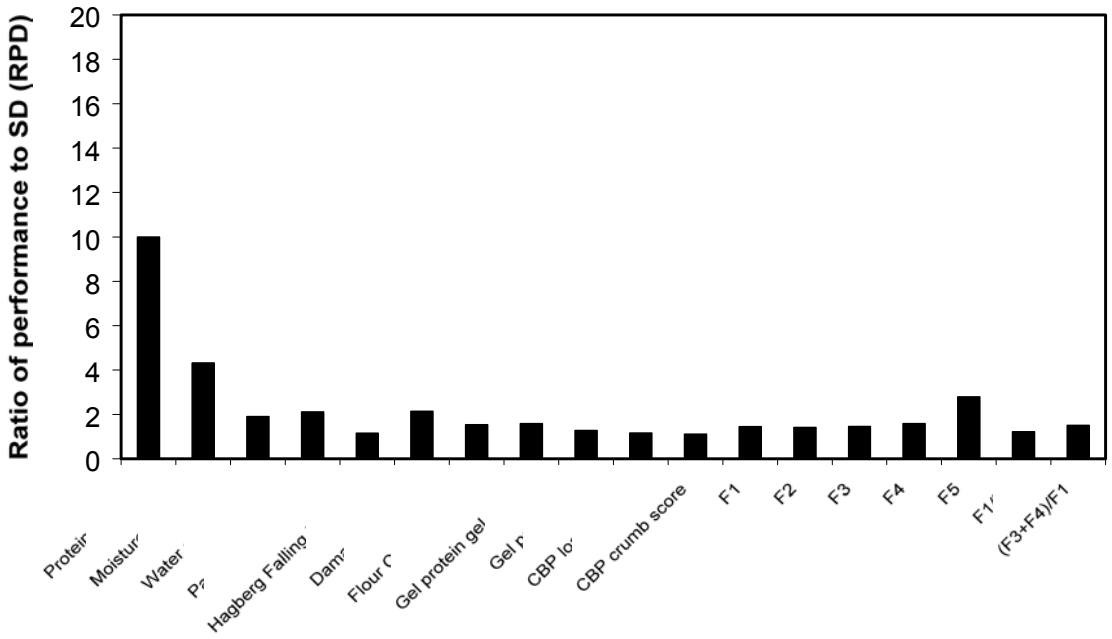
Moving from flour to ground wheat (Table 9) resulted in some deterioration in performance of the calibrations thus derived; this was probably due to greater heterogeneity of the ground wheat and also the fact that the NIR spectrum did not contain all the information about the material on which the reference analysis was performed.

Figure 7. Ratio of performance to standard deviation of reference method

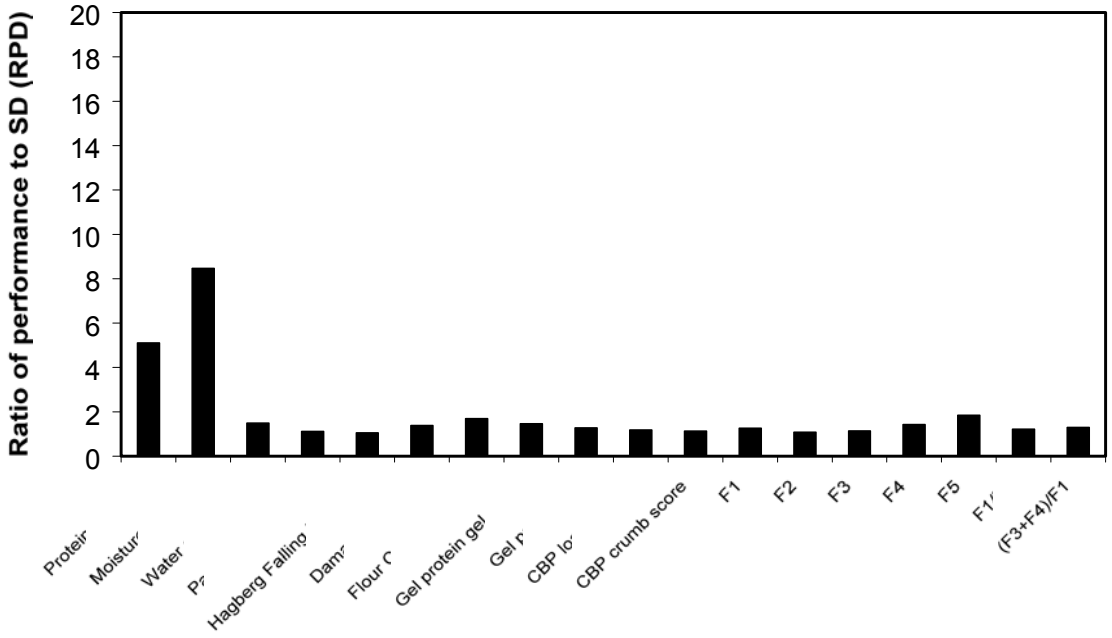
a. Flour



b. Ground wheat



c. Whole wheat



Nevertheless, the performance of the calibrations for compositional characteristics (protein and moisture content) was similar to that which has been reported previously (Millar and Scotter, 2002). The results obtained for the HPLC parameters (Figure 7b), however, were not greatly reduced from those obtained for flour; this is encouraging, indicating that an assessment of potential flour quality for breadmaking may be made at an earlier stage of processing.

The results for whole wheat were poorer than those for either of the other materials (Figure 7c), with progressive decreases in RPD being observed for all parameters except moisture content. The improved performance for this last constituent was probably due to the whole wheat being more resistant to changes in moisture status during handling than ground material, which would have a greater surface area. The whole wheat samples also had a greater range of moisture content, as demonstrated by a larger standard deviation of values (Table 10), and this will also result in larger RPD values for a given level of NIR error. This greater range of moisture contents was due to the fact that grinding using a hammer mill results in significant moisture loss and that this loss is variable.

The remaining parameters for whole grain gave poorer calibration statistics than for the other materials, however, and so confidence in the use of such calibrations for prediction of functional characteristics would be reduced. However, where validation performance was consistent, then the calibrations may still have a use where screening of many lines during a breeding programme may be required. A judgement would have to be made by the user, however, to determine whether the level of performance would be sufficient to aid with such decisions.

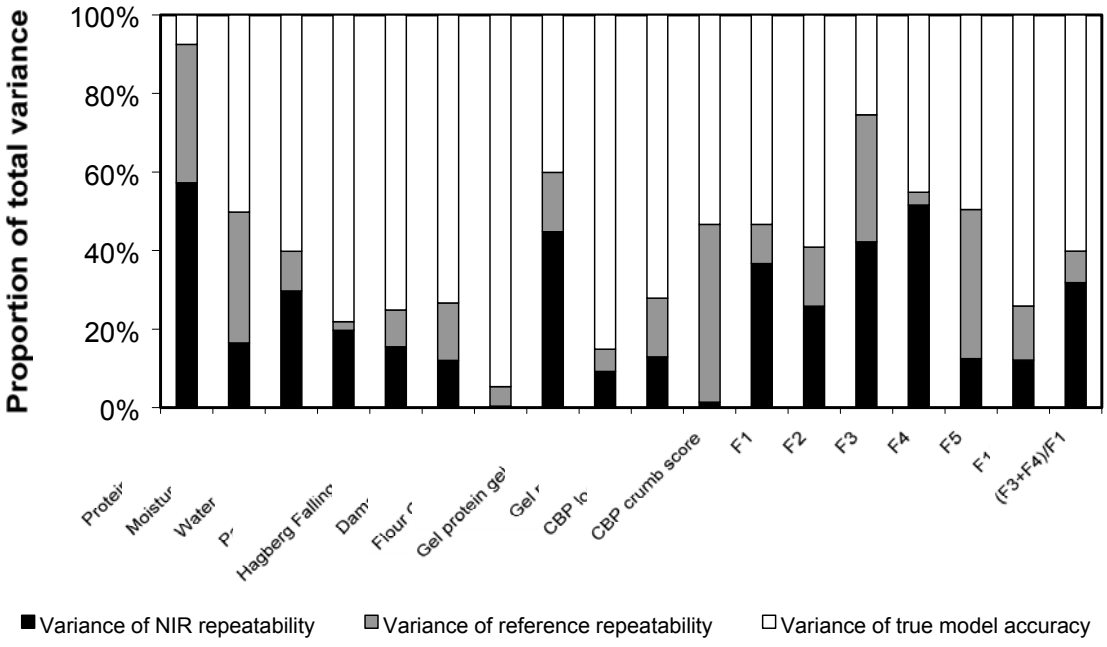
5.8 Partitioning of variance

The performance of any calibration is both a function of the precision of the methods used (NIR and the reference) and the underlying or true model error which estimates how well the datasets generated relate to one another. In this study, the relative proportions of the variances associated with each of these has been determined (Figure 8). Although such an exercise may not always help in producing better performance, it does give an indication of where the major error sources are for a given parameter. Where, for example, high errors due to poor precision were associated with the NIR determination itself, then multiple scans could be used to obtain a better estimate of the true result.

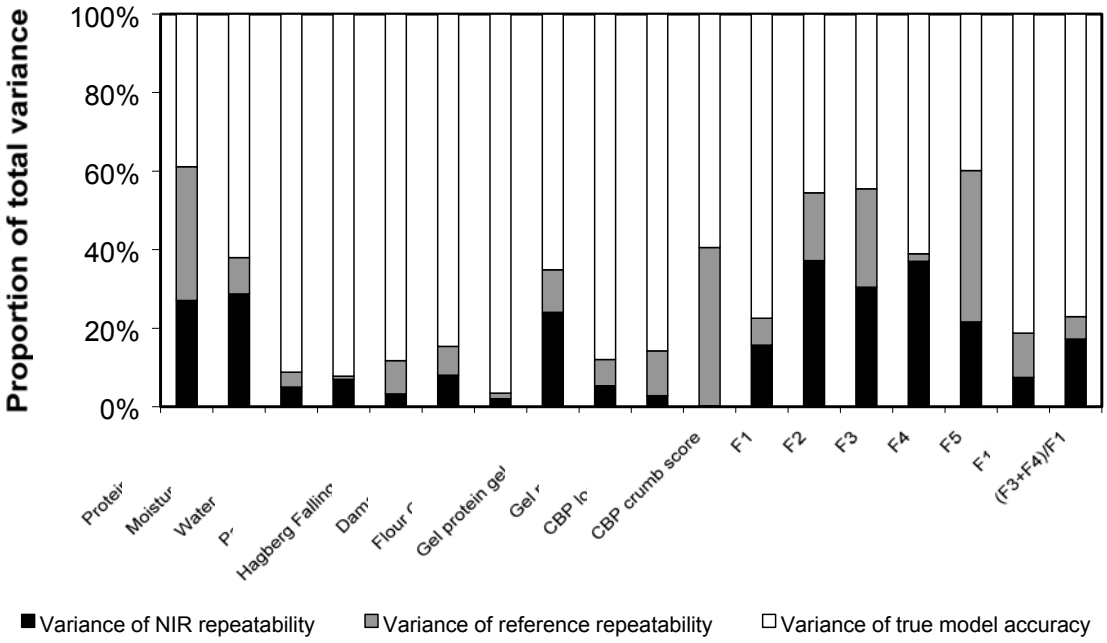
The results in Figure 8 represent the percentage of the total error associated with each of the three areas described above and care should be taken in interpreting the results as the total error (100%) in each case varies significantly with calibration performance, as demonstrated. The results for flour indicated that protein content had a very small proportion of the overall error associated with the underlying model. In other words, this calibration resulted in the NIR and combustion methods determining essentially the same thing and so a significant proportion of the underlying uncertainty of the result was due to the uncertainty of the results obtained for each of the techniques. It is unlikely, however, given the very low total errors noted for this parameter, that any further improvement in performance could be achieved through performing additional NIR scans or reference protein determinations. In any case, it would not be cost-effective to do this given the minor gains which would be achieved from the already high level of performance.

Figure 8. Partitioning of variance for materials assessed

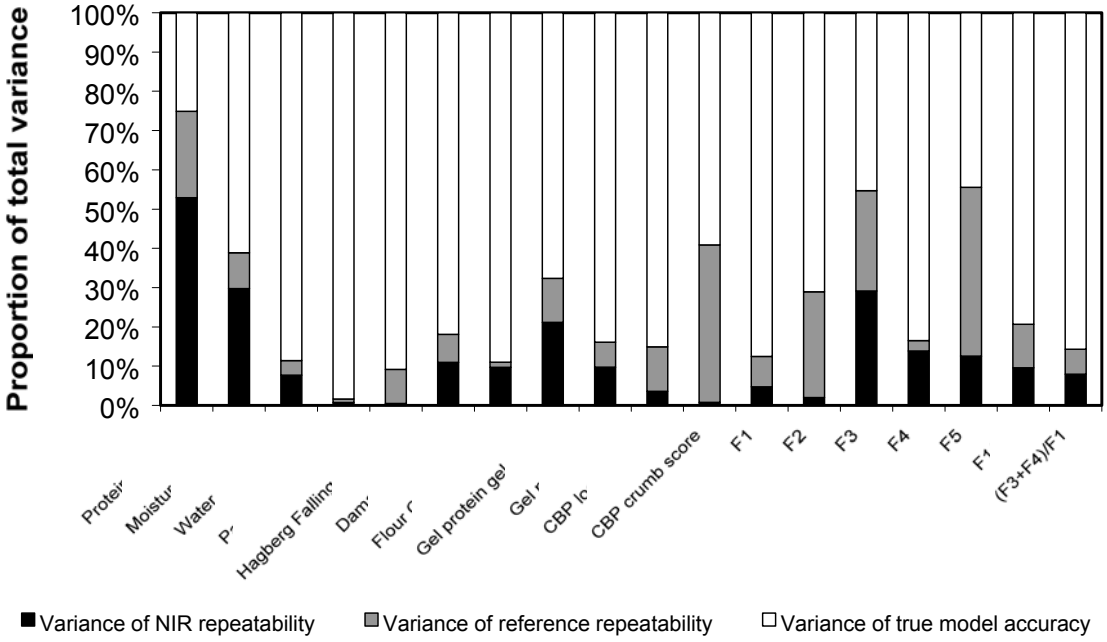
a. Flour



b. Ground wheat



c. Whole wheat



Overall, the underlying model error tended to increase with increasing complexity of the parameter to be assessed. A good example of this is gel protein G', where the NIR spectrum responds to particular constituents of the flour, while the reference method measures the rheological properties of a particular sub-portion of the flour when complexed with SDS. Here the two techniques were not measuring the same thing

and so the underlying model error was high. This is also reflected in the results for CBP loaf volume (no replication was performed for no-time doughs) where, again, the NIR spectrum will only reflect elements of flour composition, not the associated aspects of test baking. It was interesting to note the result for Flour Colour Grade, where a large proportion of the error was due to the underlying model. In this case, the overall performance was good, and this apparent anomaly results from the fact that both methods used gave very repeatable results and so, proportionally, the majority of the error was due to the calibration model. It was also interesting in this context to see that even though the overall performance of the calibrations for bread crumb score was low, the repeatability of the results when using NIR was better than that of an experienced assessor.

It might have been expected that the more heterogeneous nature of the wheat materials assessed would have resulted in higher errors in NIR repeatability. Figure 8 appears to contradict, but this is actually a function of higher overall levels of error. In some cases (such as protein content) the NIR repeatability was poorer for whole wheat than for either ground wheat or flour, as would be expected given the poorer spectral repeatability detailed above. However, this was offset significantly by two factors. In many cases, the calibrations developed for whole wheat spectra were poorer in overall performance. Calibrations having different constants (as these will have) will also result in different levels of precision, depending largely on the size of the multiplicative constants and how these interact with different parts of the spectrum (particularly the colour region where the instrument response tends to be 'noisier'). Therefore, a poor calibration could have smaller constants, leading to more repeatable results even where they were ultimately less accurate.

The second aspect is that the poorer performance of the calibrations themselves arises from higher levels of model error as the two techniques (NIR and reference) were no longer assessing the same material. This is further exacerbated for those parameters where the relationship between the material scanned by NIR and the property measured was not the same even for flour, such as test baking performance.

5.9 Partial correlations for protein functionality calibrations correcting for protein content

It has been recognised for many years that the protein content of wheat flour has a significant impact on the baking performance of that flour (Finney and Barmore, 1948). While in more recent years the focus of research in this area has rightfully been in the area of 'protein quality', the effect of increasing crude protein content on baking performance should always be taken into account. This is particularly so when assessing NIR calibrations of the type developed here. NIR may be used to predict protein content with great accuracy as has been shown right from the earliest use of the technique in the UK (Osborne, 1983) through to the results obtained in this current study. Therefore, it may be argued that where NIR appears to give good performance for calibrations related to protein functionality, this performance may be spurious due to a

significant secondary correlation with protein content, which NIR can measure with great accuracy. Indeed, in perhaps the first recorded work in which calibrations were developed for protein functionality (Osborne, 1984), it was concluded that, when the contribution of NIR's ability to measure protein content was removed, then the remaining predictive ability was negligible. Within the work reported here, therefore, two approaches have been taken to overcome this potential problem. The first of these was the sample selection procedure taken: samples were primarily chosen from those available in each harvest year on the basis of a number of properties, including protein content, test baking performance and a measure of gluten 'strength' (gel protein G'). This ensured that many of the potential cross-correlations were broken down, with samples having low protein contents but 'strong' gluten properties being included. The inclusion of samples from France at this stage was also important as the balance of gluten properties for many of these samples was quite different to those from the UK, with typically 'stronger' gluten characteristics for a given protein content.

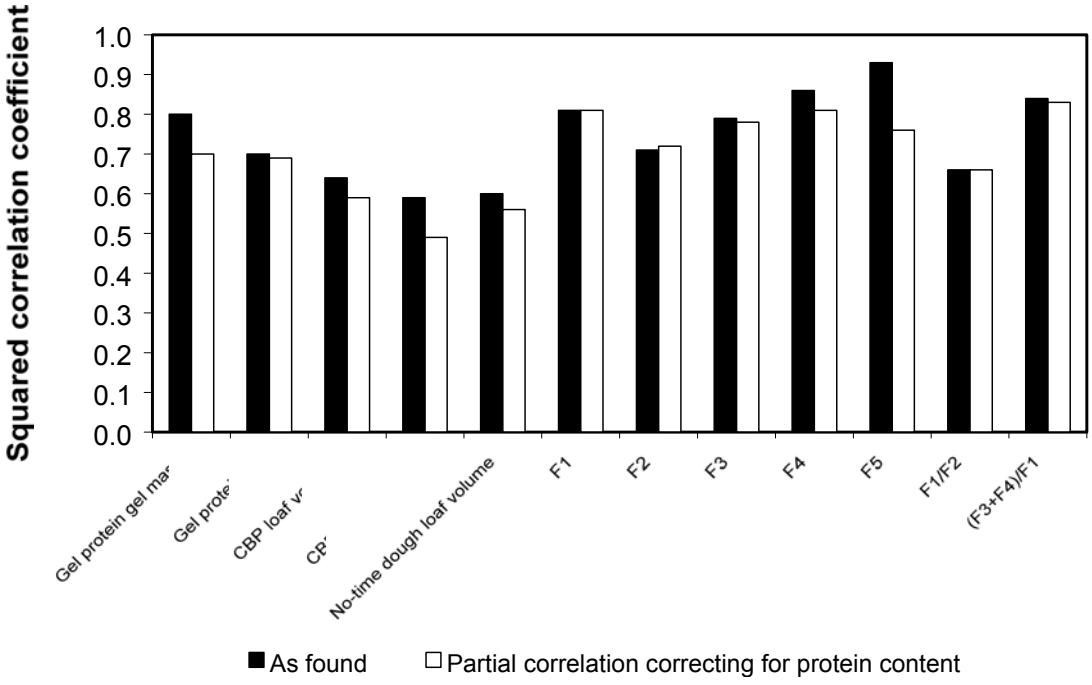
The second approach has been the use of partial correlations correcting for protein content (Fearn, 1999) as a means of assessing the worth of the calibrations following their development. This approach was used to remove the contribution of protein as a covariate from an assessment of the performance of any calibration designed to assess either protein quality or test baking performance. The technique involves an assessment of the structure of the relationship between the residuals (difference of observed from predicted values) for the NIR prediction and the true protein quality values, and the NIR prediction and the protein content for each sample. Therefore, the results may not be easily assessed using a standard error approach and so squared correlation coefficients for each treatment have been calculated (Figure 9). Where the squared correlation for a given calibration decreases when protein is removed as a covariate, the performance of that calibration depends to a greater or lesser extent on an underlying relationship with protein content. When r^2 for the partial correlation reaches 0, there is no relationship other than that for protein content.

The results indicated that, for the majority of the parameters assessed, protein content contributed in only a very limited way to the performance of the calibrations. This is a very positive result as it demonstrates that, where the global performance of the calibrations is felt to be sufficient for use (such as for (F3+F4)/F1 as a predictor of likely gluten properties), the performance of that calibration, independent of the protein content of the sample assessed, is assured. In particular, the performance of calibrations against gel protein G' and the SE-HPLC parameters F1, F2, F3, F1/F2 and (F4+F4)/F1 all gave virtually the same squared correlation coefficient when protein was removed as a covariate as when it was included. In each of these cases, therefore, the NIR calibration performance is dependent either on a direct relationship between information in the NIR spectrum and the reference method or to a secondary correlation that itself is not dependent on protein content.

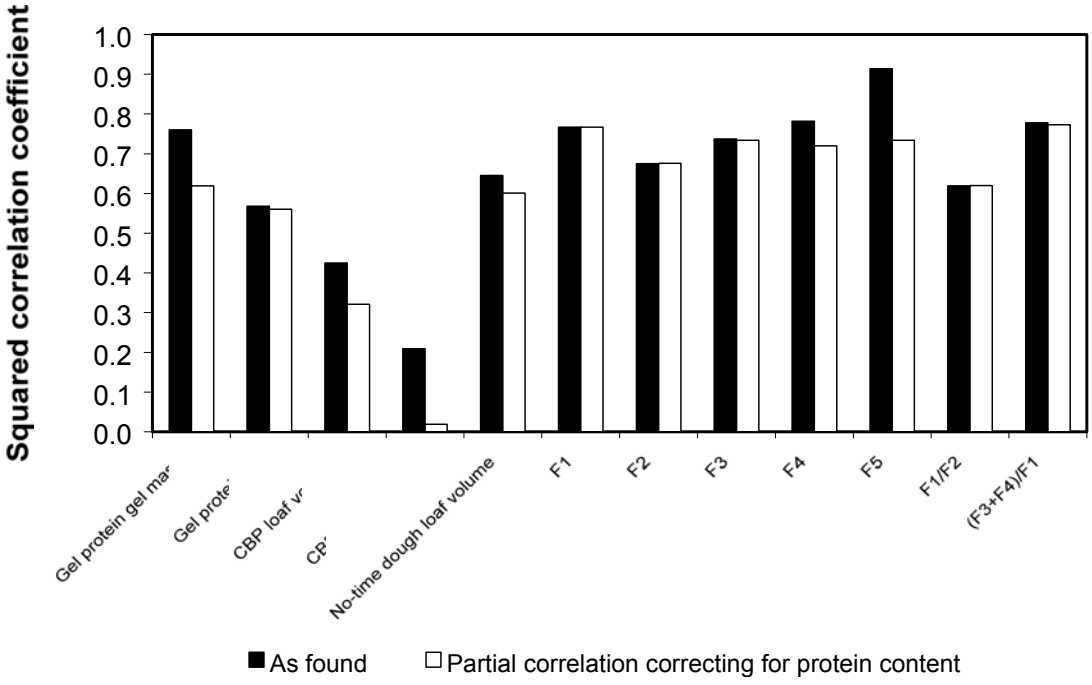
The results for gel protein mass, the three test baking attributes and the SE-HPLC parameters F4 and F5 all showed greater dependence on protein content, although only in one case (CBP crumb score for ground wheat) did this appear to be the major contributor to calibration performance.

Figure 9. Partial correlations for parameters related to protein quality

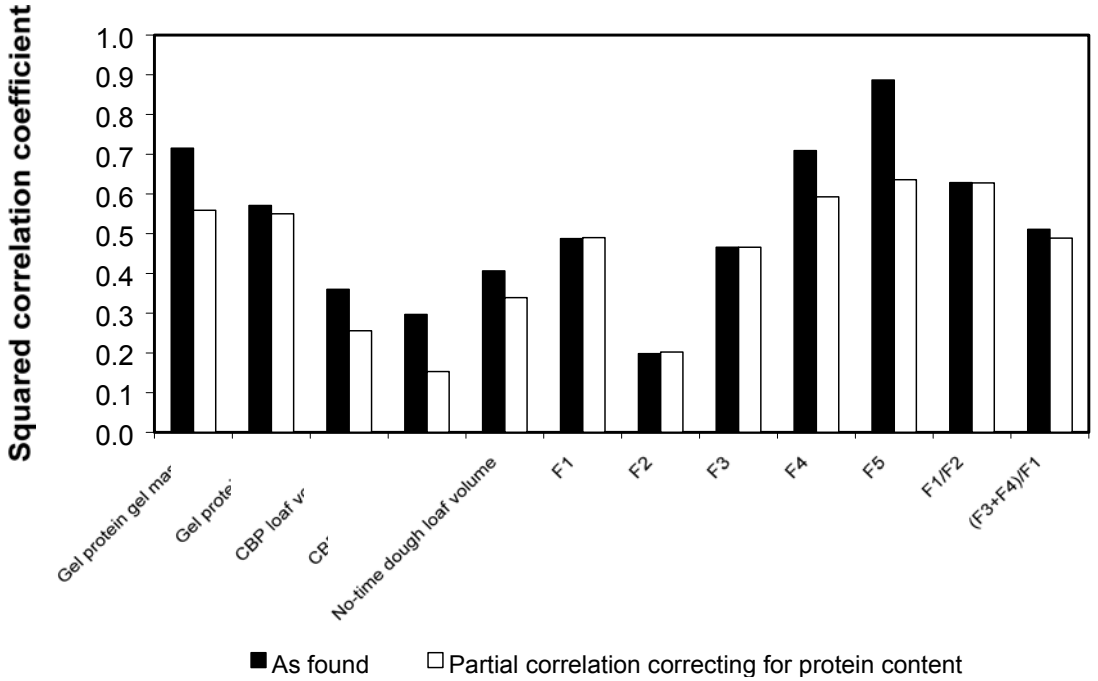
a. Flour



b. Ground wheat



c. Whole wheat



It is interesting to compare these findings with the correlations observed between the different reference results detailed in Table 3. Some caution should be taken with respect to these comparisons, as the numbers in Figure 9 emanate from r^2 while those for Table 3 are Pearsons correlation (r) and the latter will appear better as a result. Nevertheless, the comparison is of interest as the parameters which have been generally shown to have some dependence on protein content during NIR calibration development, were also those with significant positive correlations with protein content. This shows that these parameters would always be expected to have a contribution from protein content in any NIR calibration developed for them, as they inherently vary to some extent in relation to protein content. This is not surprising for a number of these parameters, including gel protein mass and loaf volume, as the relationship with protein content has been demonstrated previously by Oliver and Pritchard (1993) and Finney and Barmore (1948) respectively.

5.10 Calibration performance using separate validation samples

Although the use of cross-validation during calibration gives a more robust estimate of how the resultant calibration is likely to perform in the future, most users believe that the best means of assessing the true performance of an NIR calibration is to test using an independent set of samples. To allow a more thorough investigation of the relative performance of each of the calibrations developed, therefore, the entire dataset was split into a calibration set (3/4 of the samples) and a validation set (1/4 of the samples) with the samples being selected in order of scanning, i.e. essentially random with respect to sample properties. Calibrations were developed using the pre-treatments found to give the best performance as detailed above, except for flour F1/F2 and (F3+F4)/F1 ratios, where subsequent experience indicated that another approach was better. These calibrations were then tested using the validation set. In each case, statistical outliers were removed

using a standardised residual threshold of 2.5 for the calibration set (the default for the software used) and 3.0 for the validation set (a more conservative approach to ensure that calibration performance was not flattered by the removal of a large number of poorly-predicted samples).

The results generally indicate that the performance of the calibrations was consistent for all the various indicators of accuracy at each stage (Figures 10-12). Values for standard errors of calibration (SEC) were usually lower than those for either cross-validation (SECV) or separate validation set predictions (SEP); this is as expected given that the SEC is the most optimistic assessment of future calibration performance. In each case, the SEP was calculated as the complete error around the line of equivalence and was not corrected in any way for systematic errors (skew or bias). The regression lines in each case were plotted for the validation samples only as those for the calibration samples are constrained by the regression procedure to have a slope of approximately 1 and a bias of approximately 0. The overall error (RMSD) was used to calculate SEP to ensure that the truest possible picture of calibration performance was derived. It also served to indicate where a lack of robustness was apparent. Given the range of samples collected and analysed throughout the project, it was hoped that where calibrations were developed, their performance would be sufficiently robust to accurately predict other samples not included in the calibration set. Where this was prone to systematic error such as bias or slope deviations from 0 and 1 respectively, the performance of that calibration should be regarded with caution. Although mathematical correction for such systematic errors may be applied, it is often an indication of instability and so an adjustment on one occasion may be followed by the apparent need to adjust in the opposite direction in a subsequent case. It is the author's view that such adjustment is undoubtedly warranted when correcting the performance of one instrument to that of another, but that when assessing calibrations on the same instrument as that on which calibrations were developed, such deviations in performance would not be expected for calibrations likely to perform well in the future.

The performance of protein content calibrations stands out for all the materials as being particularly good. This was especially marked for flour, where the trends observed for the calibration work were repeated for the validation set, with prediction errors lower than any reported before to the author's knowledge. The performance of the calibrations for moisture content were also very good and were well within the limits of performance used by the UK milling industry (ICC, 1986). The results for these two 'compositional' parameters also serve to indicate the absolute performance of NIR calibrations as well as the underlying soundness of the data collected during this study.

Given the rather more complex nature of the relationships between NIR spectra and the other measured parameters, it is not surprising to note that in absolute terms the performance appeared less accurate for them. Although the error figures for different parameters are not directly comparable, due to the differences in measurement scale, an overall impression of performance may be gained by examination of the graphs included (Figures 10-12) and by reference to the repeatability of the reference methods used as detailed

above. The results tend to fall into two broad categories: those which have some potential for use in screening type applications, and those which demonstrate performance which is likely to be insufficient for practical use. For flour, Hagberg Falling Number, the three test baking parameters and SE-HPLC F2 all fall into the latter category, with gel protein G' and SE-HPLC F1/F2 being borderline cases. The poor performance for these calibrations not only tends to manifest itself in generally high errors when compared with the standard deviation of the reference values obtained, but also tends to exhibit systematic errors for the validation set, particularly as skew, which leads to the conclusion that the performance of these calibrations is likely to be unstable.

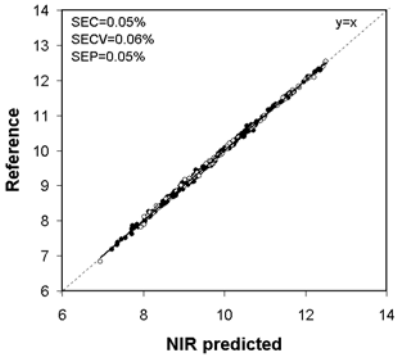
Generally those calibrations which give poor performance are those which might be expected to do so for a variety of reasons. In the case of Hagberg Falling Number, the particular enzyme of interest, *alpha*-amylase, is but one of many proteins naturally found in wheat. The chances, therefore, of determining its concentration with any degree of certainty with a technique such as NIR, which primarily responds to N-H and C=O bonds in the amide groups of all proteins, are very low. Any success with such a parameter will almost certainly be as a result of a secondary correlation with related characteristics, such as particle size. The sample selection procedure adopted in this case as the advantage of reducing such secondary correlations such that, while calibration statistics in some cases may appear poorer than more generic studies, the estimation of likely true performance is probably more realistic.

For the parameters associated with test baking, the poor performance was due to the lack of a good underlying model and not due to poor repeatability of the reference methods (although this is certainly poorer than for other parameters) as shown above. This is because of the processing involved (the effects of which will not be included in the NIR spectrum of wheat and flour), which interacts with flour quality (which may be related to the spectral data) to produce a quality which is not easy to predict, even using the reference techniques. The difficulty of predicting such characteristics has previously been demonstrated by a number of workers (Osborne, 1984; Delwiche and Weaver, 1994). The results of this latter study were similar to those found here with the prediction of bread cell structure being highlighted as particularly difficult.

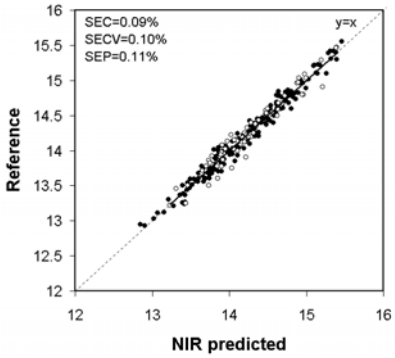
It is not clear why the performance for SE-HPLC F2 was the poorest of all the SE-HPLC parameters, although it is clear that this has a significant effect, in turn, when attempting to predict the associated ratio F1/F2 directly. The remaining calibrations all appeared to give reasonably consistent performance with little indication of systematic errors. It is interesting to note, however, that all of the parameters which are thought to relate, in particular, to the molecular weight distribution of the polymeric proteins, i.e. gel protein G', SE-HPLC F1 and F1/F2, appear to consistently under-predict the samples at the top end of the range assessed (Figure 10). This has been noted in other exercises of this type carried out by the team at ARVALIS (Christine Bar-L'Helgouac'h, personal communication).

Figure 10. Calibration and validation performance for flour (closed symbols represent calibration set, open symbols the validation set)

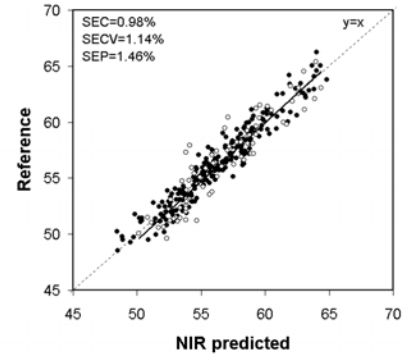
Protein content as is (%), N × 5.7



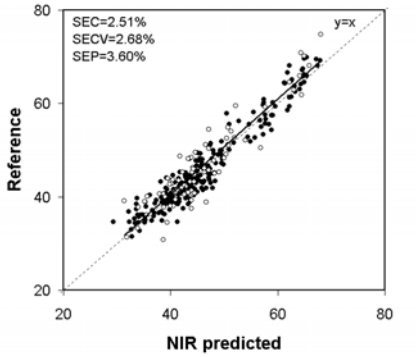
Moisture content as is (%)



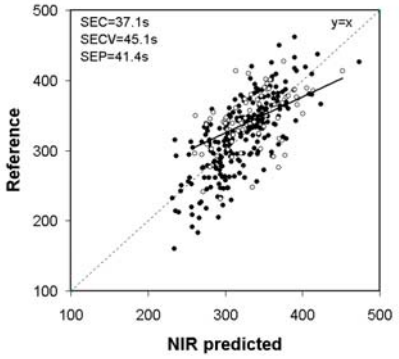
Farinograph water absorption as is (%)



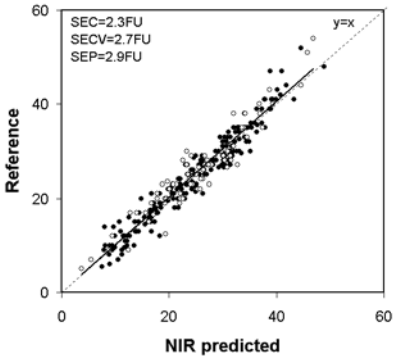
Particle Size Index (%)



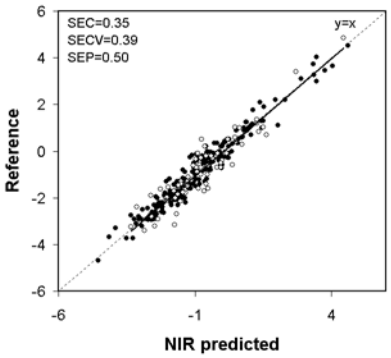
Hagberg Falling Number (s)



Damaged starch (FU)



Flour Colour Grade



Gel protein mass (g)

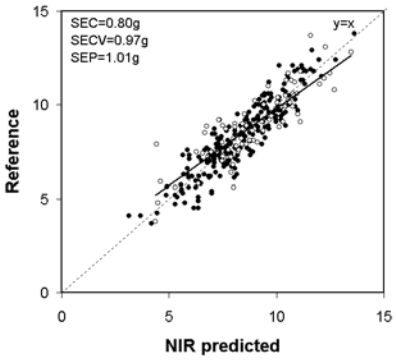
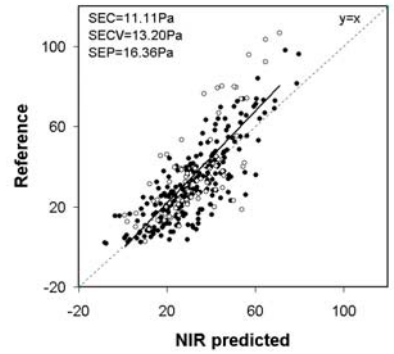
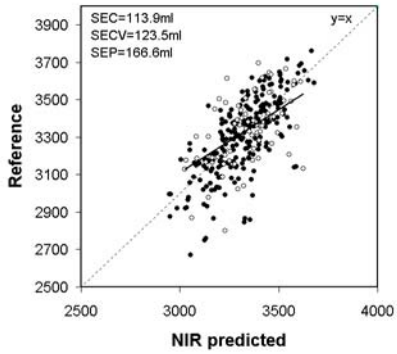


Figure 10. Calibration and validation performance for flour (closed symbols represent calibration set, open symbols the validation set)

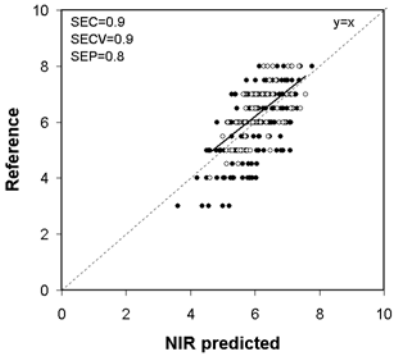
Gel protein G' (Pa)



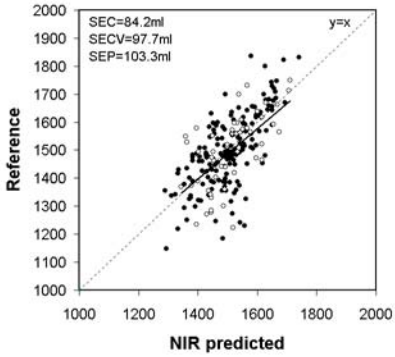
CBP loaf volume (ml)



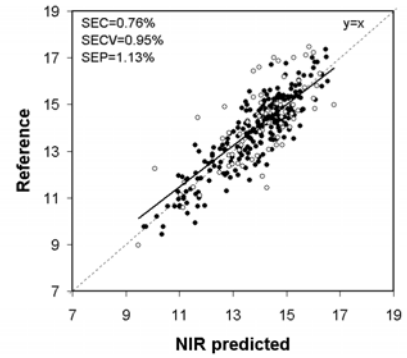
CBP crumb score



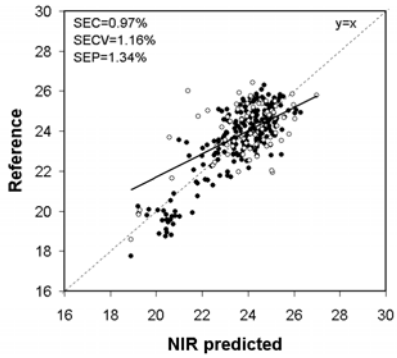
No-time dough loaf volume (ml)



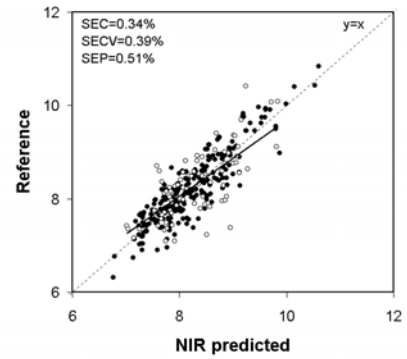
SE-HPLC F1 (%)



SE-HPLC F2 (%)



SE-HPLC F3 (%)



SE-HPLC F4 (%)

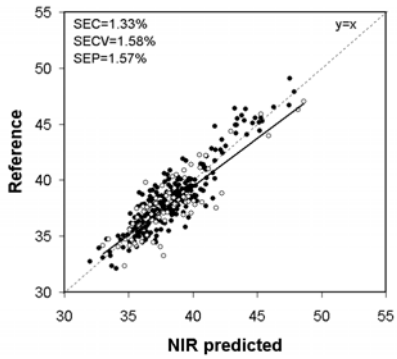
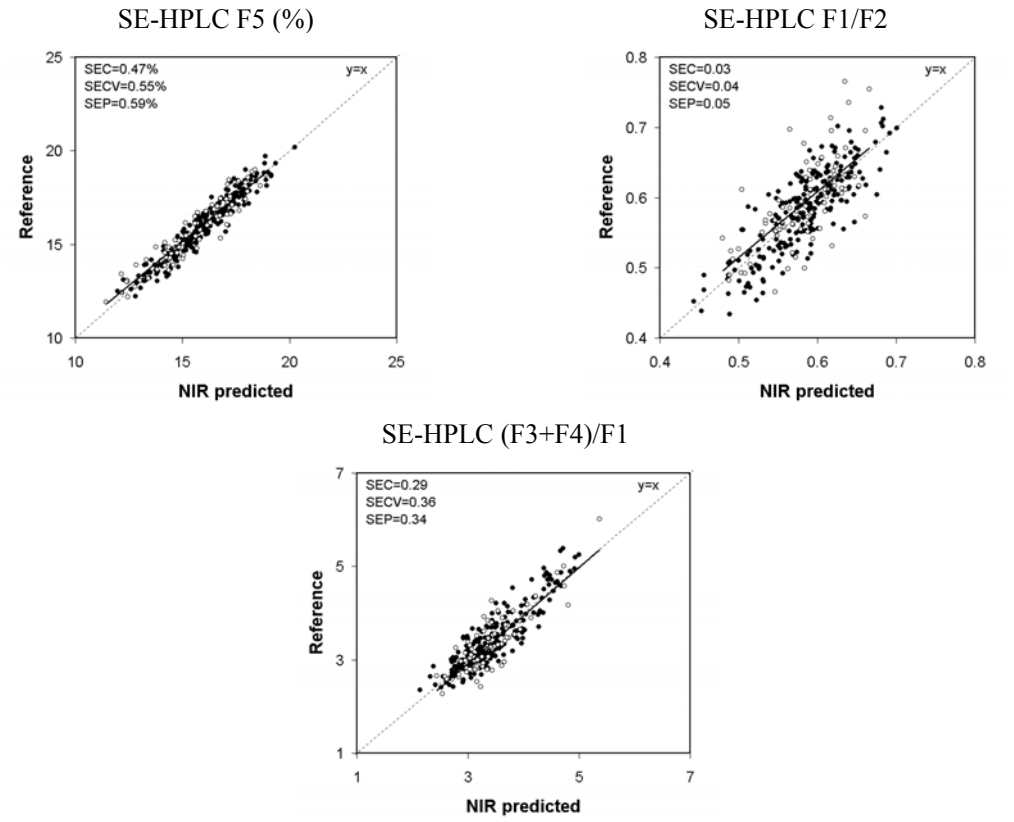


Figure 10. Calibration and validation performance for flour (closed symbols represent calibration set, open symbols the validation set)



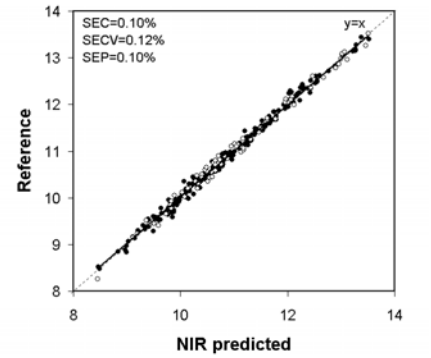
This indicates that fundamentally the two techniques (the reference method and NIR) do not measure the same entity. Rather, the information in the NIR spectrum relates to another element of wheat quality which relates well to protein molecular weight distribution for much of the range, but fails to model the variation exactly, particularly at the higher end of the scale of interest. The results above indicate that this correlation is not simply based on protein content and so clearly is more complex and may relate in a more consistent way to protein quality and ultimately baking potential.

Overall, therefore, it may be concluded that at least reasonable performance for the majority of the NIR calibrations developed has been demonstrated. In terms of recommendations for use as a means of assessing protein quality, all of the SE-HPLC parameters other than F2 and F1/F2 may be used with some confidence and, in particular, (F3+F4)/F1 looks to have good potential for use as a screening tool to segregate flour of breadmaking quality from that more suitable for lower premium uses.

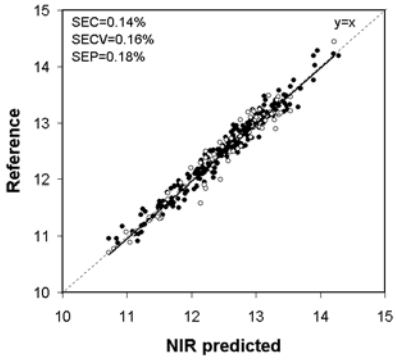
Although other workers have not published statistics for calibrations developed using the same SE-HPLC protocols, general comparisons may be made with other work which was similar in concept (Wesley *et al.*, 1999) to that detailed here.

Figure 11. Calibration and validation performance for ground wheat (closed symbols represent calibration set, open symbols the validation set)

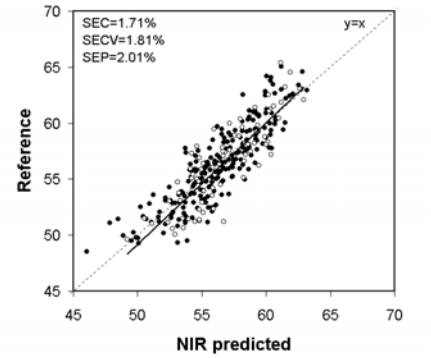
Protein content as is (%), N × 5.7



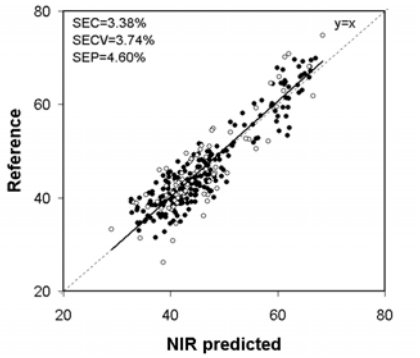
Moisture content as is (%)



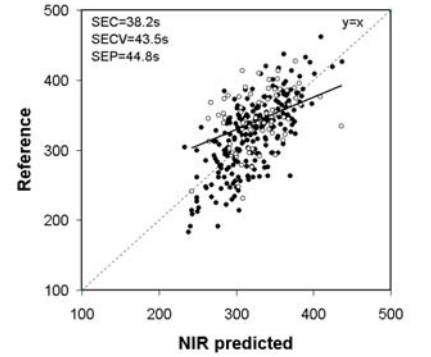
Farinograph water absorption as is (%)



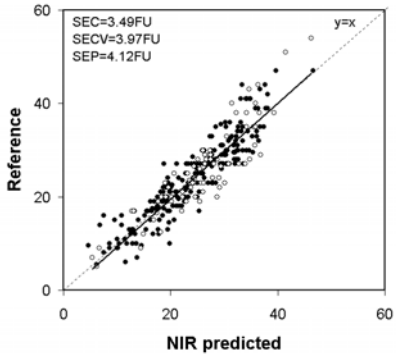
Particle Size Index (%)



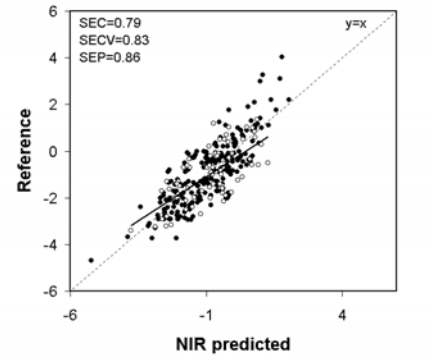
Hagberg Falling Number (s)



Damaged starch (FU)



Flour Colour Grade



Gel protein mass (g)

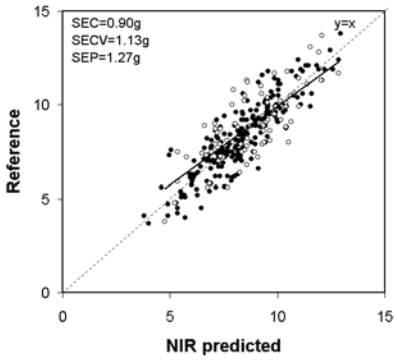
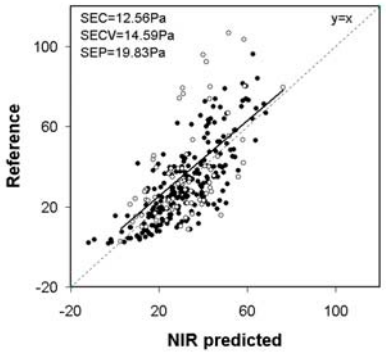
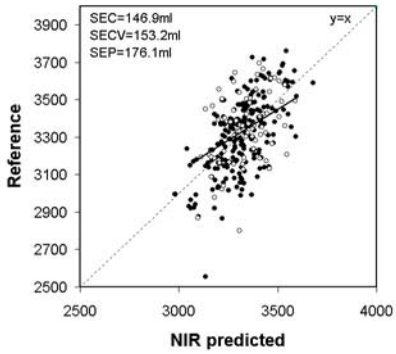


Figure 11. Calibration and validation performance for ground wheat (closed symbols represent calibration set, open symbols the validation set)

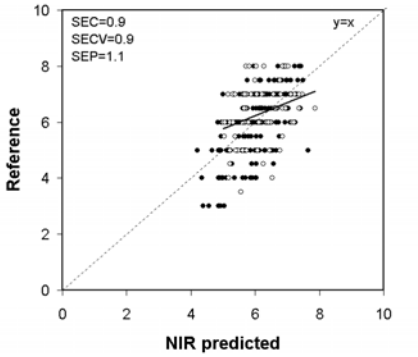
Gel protein G' (Pa)



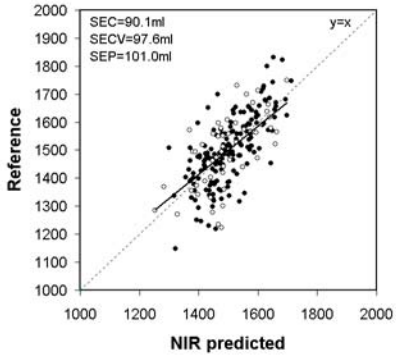
CBP loaf volume (ml)



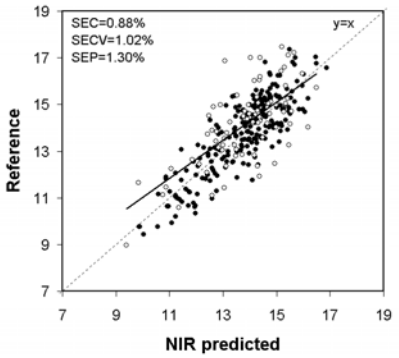
CBP crumb score



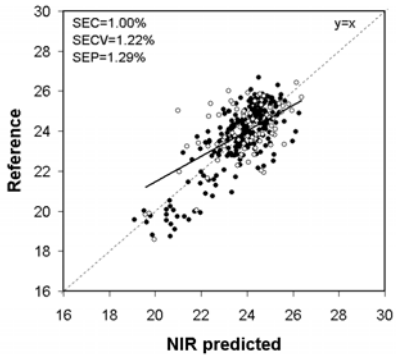
No-time dough loaf volume (ml)



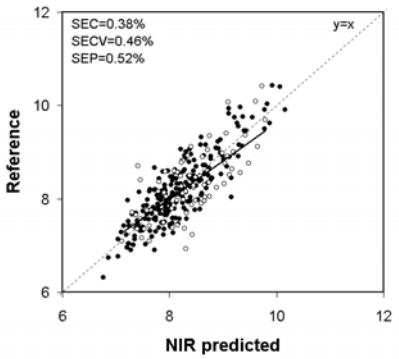
SE-HPLC F1 (%)



SE-HPLC F2 (%)



SE-HPLC F3 (%)



SE-HPLC F4 (%)

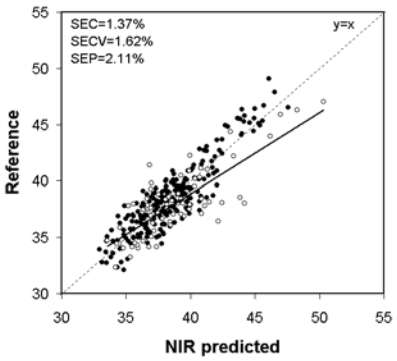
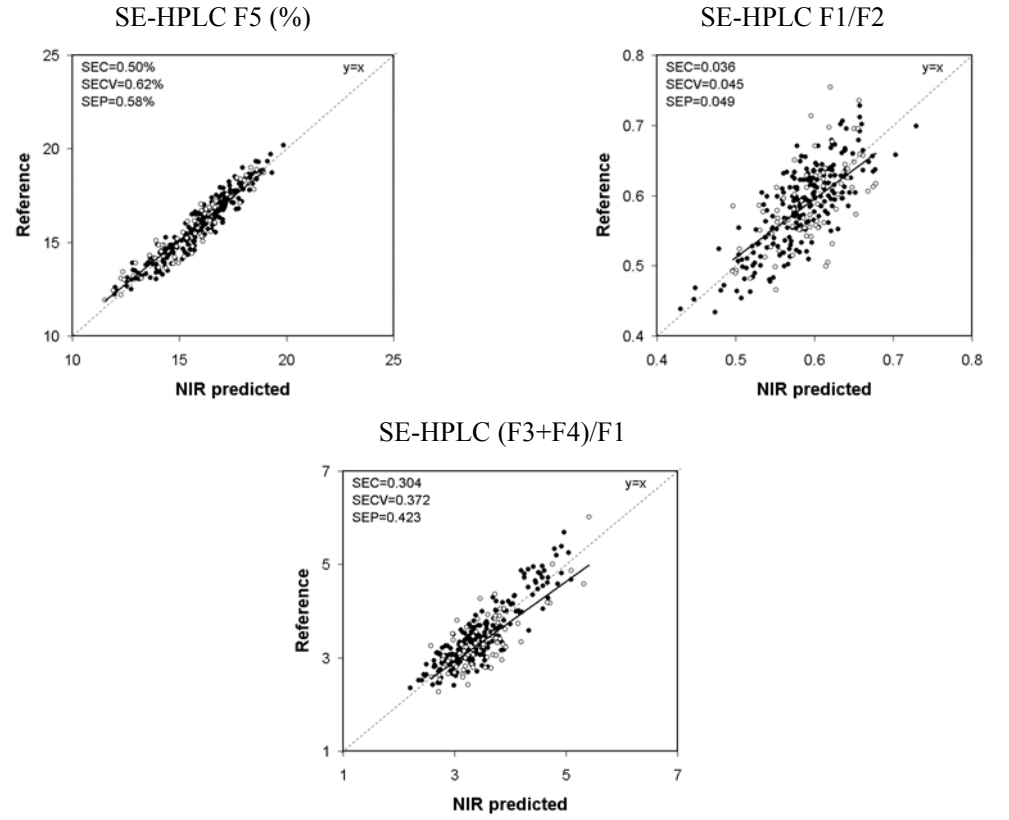


Figure 11. Calibration and validation performance for ground wheat (closed symbols represent calibration set, open symbols the validation set)

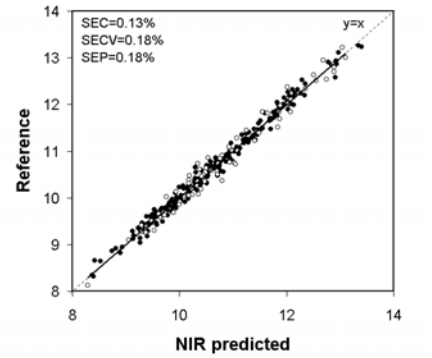


Their results for squared correlations of glutenin and gliadin contents of flours were similar to the results from this study, although their approach differed in two ways. The first was that the SE-HPLC analysis resulted in discrimination of fewer fractions and would be expected to be less sensitive to minor changes in protein functionality. Secondly, the 'calibration' approach adopted by Wesley *et al.* was based on fitting individual NIR spectra for gluten, gliadin and starch rich materials to allow prediction of these components in other systems. The advantage in such an approach is the protein content-independent nature of the calibrations thus developed. This is offset, however, by the more complex and less routine calibration approach. The current study has resulted in calibrations of at least equivalent performance but the range of protein characteristics for which calibrations have been developed and the calibration approach taken should both ensure greater ease of use and more available information.

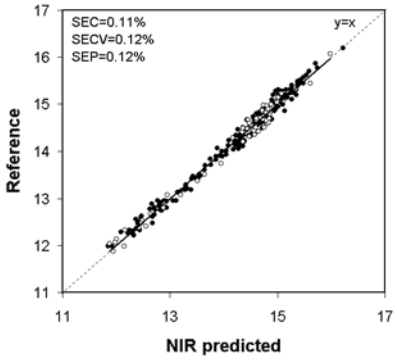
The calibrations developed for ground wheat and whole wheat (Figures 11 and 12) gave similar trends in performance in the majority of cases, although the overall results were generally poorer for the less transformed material. It was interesting to note, however, the results for loaf volume when produced using a no-time dough approach. In this case, the results for ground wheat were very similar to those obtained for flour, albeit at a level for both which would not give a user a great deal of confidence when quoting results.

Figure 12. Calibration and validation performance for whole wheat (closed symbols represent calibration set, open symbols the validation set)

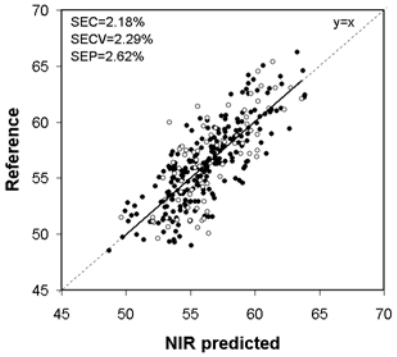
Protein content as is (%), N × 5.7



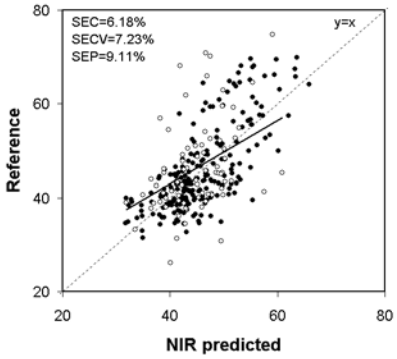
Moisture content as is (%)



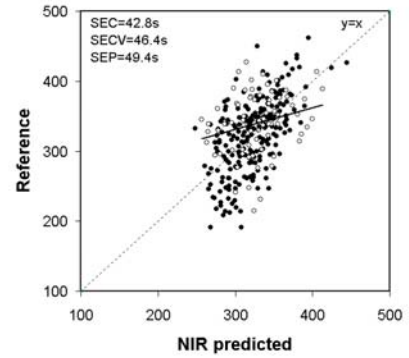
Farinograph water absorption as is (%)



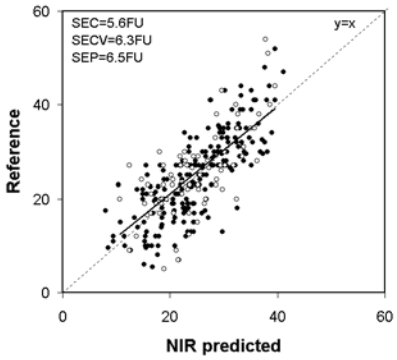
Particle Size Index (%)



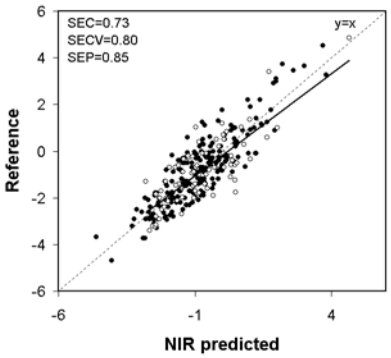
Hagberg Falling Number (s)



Damaged starch (FU)



Flour Colour Grade



Gel protein mass (g)

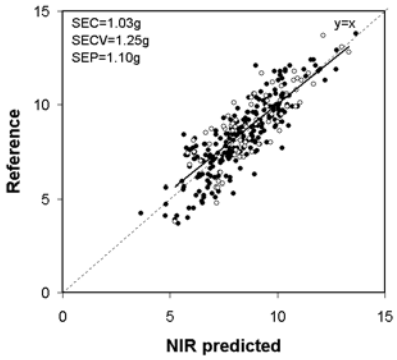
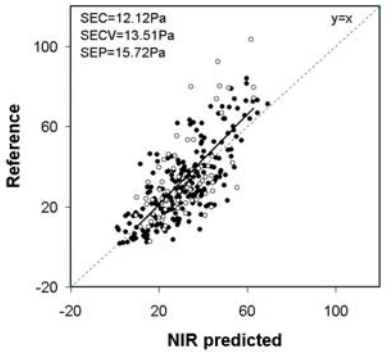
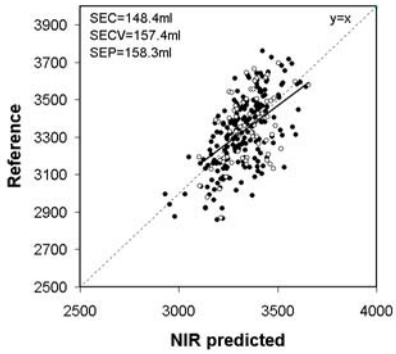


Figure 12. Calibration and validation performance for whole wheat (closed symbols represent calibration set, open symbols the validation set)

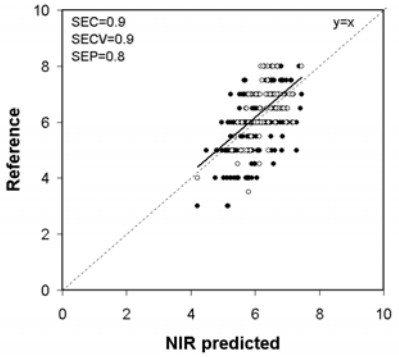
Gel protein G' (Pa)



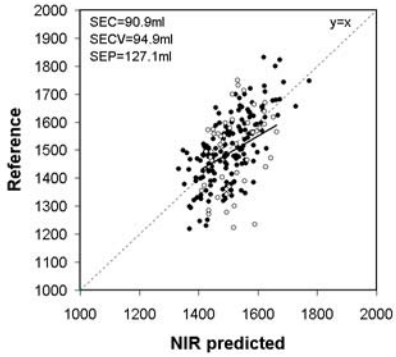
CBP loaf volume (ml)



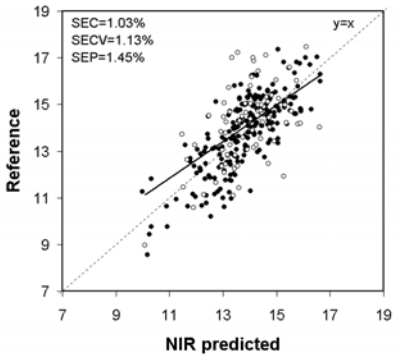
CBP crumb score



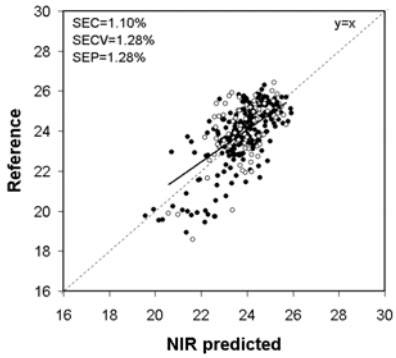
No-time dough loaf volume (ml)



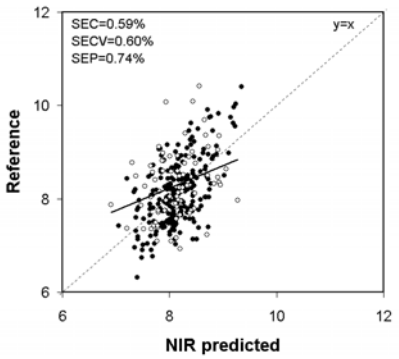
SE-HPLC F1 (%)



SE-HPLC F2 (%)



SE-HPLC F3 (%)



SE-HPLC F4 (%)

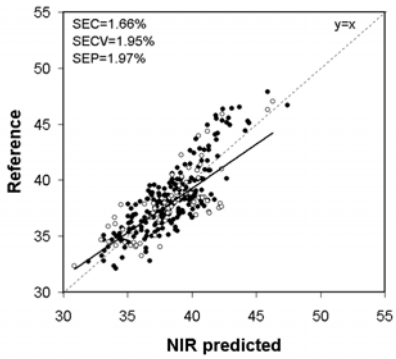
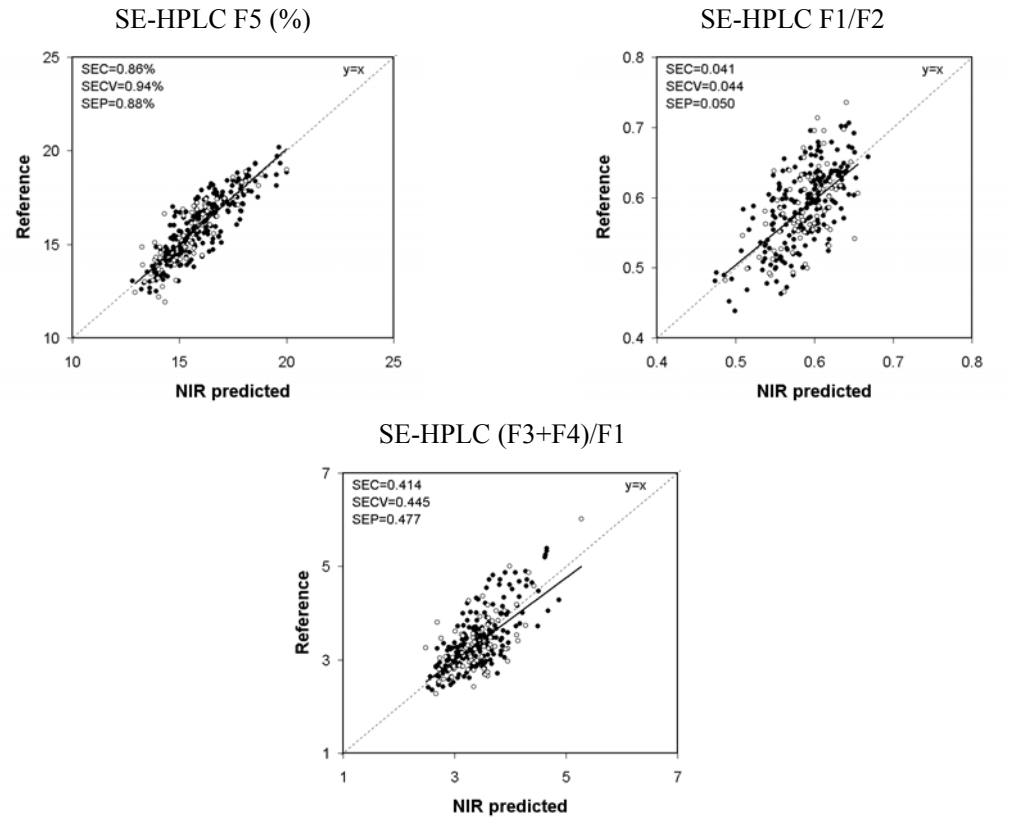


Figure 12. Calibration and validation performance for whole wheat (closed symbols represent calibration set, open symbols the validation set)



The results obtained when using spectra collected from whole wheat were considerably poorer, indicating virtually no true predictive potential. This trend was reversed when CBP loaf volume was assessed, which serves to indicate that as the performance levels decrease, the consistency of performance for different materials or parameters also becomes less.

The main difference observed for the different materials was for parameters related to wheat endosperm texture, where calibrations based on whole wheat spectra gave much poorer performance. The most obvious example of this was for particle size index. This would be expected, however, given the recognised sensitivity of NIR to particle size. It is not that surprising to find, therefore, that the results for whole wheat for this parameter were very poor. However, recently, workers have been attempting to develop calibrations relating hardness to assessment of whole grain spectra. The results obtained here would tend to indicate that the accuracy of such calibrations is likely to be significantly less than that currently accepted for use with ground wheat or flour. In turn, the calibrations for damaged starch and, to a lesser extent, water absorption were poorer for the whole grain material, due to the way in which both these parameters are thought to be related to wheat endosperm texture (Farrand, 1969).

5.11 Spectral basis for NIR calibrations developed

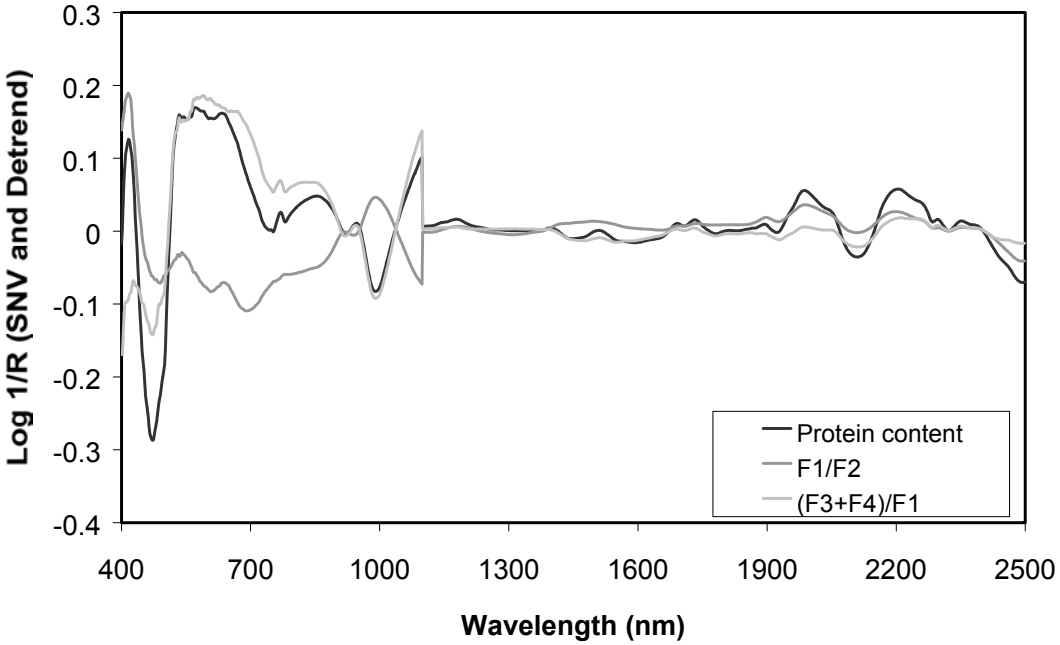
Typically NIR is used as a technique to determine physical, compositional or functional parameters of materials, with little importance being assigned to the underlying spectroscopy involved. This is not necessarily incorrect or indicative of poor practice, as the goal of all calibration exercises should be to offer accessible solutions to the end-user. Nevertheless, at some stage during calibration development, an attempt to interpret the basis for the performance of the calibration is important to ensure that robust equations result. In recent years, there has been renewed interest in the use of NIR as a pure spectroscopic tool to probe molecular interactions, particularly in the cereals area where a number of studies have highlighted its application to the assessment of the changes occurring during dough mixing (Wesley *et al.*, 1998b; Millar *et al.*, 1999).

All of the calibrations developed using the full spectral range of the Foss NIRSystems 6500 instrument were produced by modified partial least squares regression (MPLSR). This technique generally produces more efficient solutions to multivariate problems than other approaches such as principal components regression (PCR). This is because variation in both the multivariate spectral response and the univariate (PLS1 type) or multivariate (PLS2 type) reference data is used to develop the most predictive factors, rather than those components which best described the spectral data alone as with PCR. The major drawback of such an approach, however, is that the inclusion of a contribution from the reference data makes the interpretation of the spectral loadings difficult. To overcome this, the results for protein content and the two SE-HPLC ratios, F1/F2 and (F3+F4)/F1, were grouped such that the flour spectra representing the top 5 measured values for each parameter had the flour spectra representing the bottom 5 subtracted. In each case the spectra were transformed using Standard Normal Variate and Detrending operations to reduce the effect of particle size or pathlength variation on the results. The difference spectra are presented in Figure 13.

The results show that a number of features were apparent in both the visible/near visible (400-1098nm) and NIR (1100-2498) regions. In the former, interpretation of the data is difficult as this region is primarily associated with the electronic transitions associated with the absorption of visible light and because no previous studies have attempted to assess the spectral variation in this region for such parameters. However, it may be concluded that the spectral differences for (F3+F4)/F1 and protein content were similar in this wavelength range and that little discrimination of these parameters on this basis may be effected. The results for F1/F2 were quite different, however, with significantly lower absorbance in the visible region from 500-700nm. Lower absorbance in this region would equate to greater reflectance, which would indicate a brighter appearance, and this may demonstrate an underlying relationship for this dataset between flour colour and protein functionality. It is interesting, therefore, to refer to the results obtained when attempting to predict CBP loaf volume from the other reference data. Where Flour Colour Grade was included in the

equation it had a negative constant, indicating that brighter flour colour was positively correlated with test baking performance.

Figure 13. Flour difference spectra for protein content F1/F2 and (F3+F4)/F1



The most notable difference, apart from that in the colour region, was around 1000nm, where the results for F1/F2 were the opposite to those for the other two parameters assessed. This wavelength region has previously been assigned to the second overtone of O-H stretch (Osborne *et al.*, 1993) and it is not immediately clear why this should be significant. The region around 1020nm has been assigned to protein, however, and it may be that this is the cause of the difference observed.

In the NIR region, all parameters were characterised by peaks in the regions around 1180, 1514, 1694, 1734, 1992, 2050, 2208, 2302, 2356 and 2388nm and troughs around the regions 1926, 2114, 2290 and 2326nm. These wavelengths are virtually identical with those found in a similar study by Delwiche *et al.* (1998). This is encouraging as it indicates that the calibrations developed in this study were not simply specifically adapted to the samples included in this study but would appear to rely on spectral characteristics which are of a more universal nature. It is interesting that the spectral difference plots were very similar for all the parameters and at first acquaintance slightly confusing, given that the results for the partial correlations were so clear. The gluten forming proteins are the main components of the overall protein content of wheat flour. It is entirely logical, therefore, that the spectral patterns associated with various fractions of them would be largely similar to that for protein content, albeit with specific differences in accordance with their different structures and functions. In this case, these differences exist in the regions around 1900 for protein content

and F1/F2 when compared with (F3+F4)/F1, and 1552, 1714, 1784 and 2258nm for protein content and (F3+F4)/F1 when compared with F1/F2. Clearly, it is the relative absorbances in each of these regions which is responsible for the performance of the NIR calibrations derived in this study. The conclusion of the study by Delwiche *et al.* (1998) was that the region 1850-2350, primarily consisting of combination bands of CH, OH and NH, was the main difference related to gluten protein quality and this region also forms the basis for at least two of the regions of interest.

Three of the regions noted above (2050, 2306 and 2350nm) were also found to be of interest in differentiating the spectra of 'purified' samples of gliadin and glutenin (Wesley *et al.,* 1999). The first of these was assigned to amide groups; there is some rationale for this given the recognised high proportion of glutamine residues in high molecular weight glutenins (Belton, 1999) and the previously reported interest in the assessment of amide content as a means of assessing flour quality (MacRitchie, 1979; Ewart, 1982). However, the results obtained in this case indicate clearly that this wavelength was a feature (albeit in varying degrees) for all the parameters studied, including protein. As the relative contribution was greatest for the (F3+F4)/F1 ratio, it may be concluded that the relationship between this wavelength and protein quality is not as clear as has been previously reported and that other regions may offer better discriminative potential. The regions around the wavelengths at 2306 and 2350nm were also found in this study but again were linked to all the parameters. Previous assignment of these had implicated lipid inclusion in the glutenin fraction of wheat flour (Wesley *et al.*, 1999), corroborating findings by other workers (Pawlinsky and Williams, 1998). Again, it must be concluded from this study that the discriminative potential of this spectralregion for real world samples on the basis of protein quality is limited and that calibrations based on the other regions already highlighted offer better potential.

5.12 Calibrations developed using a filter instrument approach

The calibrations developed in the work reported up to this point were all developed using the full range collected using the Foss NIRSystems instrument in the NIR region and, on occasions, the visible, near visible and NIR regions. Such an approach is the most flexible as it is not limited by the unavailability of particular wavelength regions and allows the use of extensive data pre-treatments options. However, the majority of NIR instruments used by the UK grain trade are based on filter rather than full spectrum technologies, which means that a relatively small number of discrete wavelength regions are available rather than the entire range. This is not necessarily a problem for the mainstream applications such as protein or moisture contents in cereals and their products, where the spectral features to be assessed are relatively large. However, iIt is more of an issue when relatively subtle parameters are assessed and it is interesting in this context that the feed industry tends to employ more sophisticated instruments of the type used here due to the complexity of the systems to be assessed and the large number of trace materials which need to be analysed.

In an attempt to take account of the large number of filter instruments used in the UK grain trade, however, calibrations were developed for wavelengths typical of those found in a number of different types of commercially-available filter instruments. The results for this exercise are given in Table 11. They demonstrate very clearly the potential or otherwise of a filter instrument approach for the assessment of the parameters included in this study. In common with the scanning instrument, and as would be expected given previous experience (Osborne *et al.*, 1982b), the calibrations developed for protein and moisture contents, water absorption, Particle Size Index, damaged starch and Flour Colour Grade all indicated reasonable performance. That for SE-HPLC fraction F5 also had some potential, possibly as a result of the negative correlation with protein content for this parameter as shown above. However, the remaining parameters gave such poor performance as assessed by R^2 and SECV results that no recommendation of them for routine use may be made. Given the subtle variations seen in the spectral difference exercise for SE-HPLC parameters F1/F2 and (F3+F4)/F1 seen above, it is not surprising to see that individual 'slices' of information based on individual wavelength regions cannot capture the same amount of detail as may be seen for the full spectrum, where slight variations in peak positions and relative intensities were apparent. Given the quality of the data collected, the range and number of samples and the relative difficulty of calibration development for full scan data, it can only be concluded, therefore, that more sophisticated instrumentation (such as that used in this study) is required to allow acceptable performance to be achieved for the calibrations derived. While the financial investment involved in such instrument types is greater than that for those based on filter technology, it may be argued that the benefits in terms of the range of the other parameters which may be assessed and the accuracy with which all the parameters may be estimated would out-weigh the financial considerations. However, such considerations need to be undertaken by the ultimate end-users of the work described in this report and are outside the scope of this discussion. However, for those who currently use laboratory grade NIR spectrometers, the calibrations developed here may now be made available for their use.

5.13 Development of calibrations using combined datasets

The two main criteria for ensuring that NIR calibrations will perform well in practice are that the calibrations should give accurate results and that they should be robust, i.e. work for a wide range of samples. To ensure that the latter was the case for the calibrations developed through this work, a programme of exchange of whole wheat samples was undertaken. These samples were then milled, analysed and test baked following typical UK practice (for French-grown wheat) and French methods (for the UK-grown wheat). In addition, the information emanating from this exercise was thought to be of potential interest for those sourcing wheat from the global market. The laboratory milling procedures were different (Quadrumat Senior in France and Bühler in the UK) and this will have resulted in flours having different properties. Such variation would be encountered in 'real-world' situations, however, and so calibration performance for the range of samples produced was evaluated.

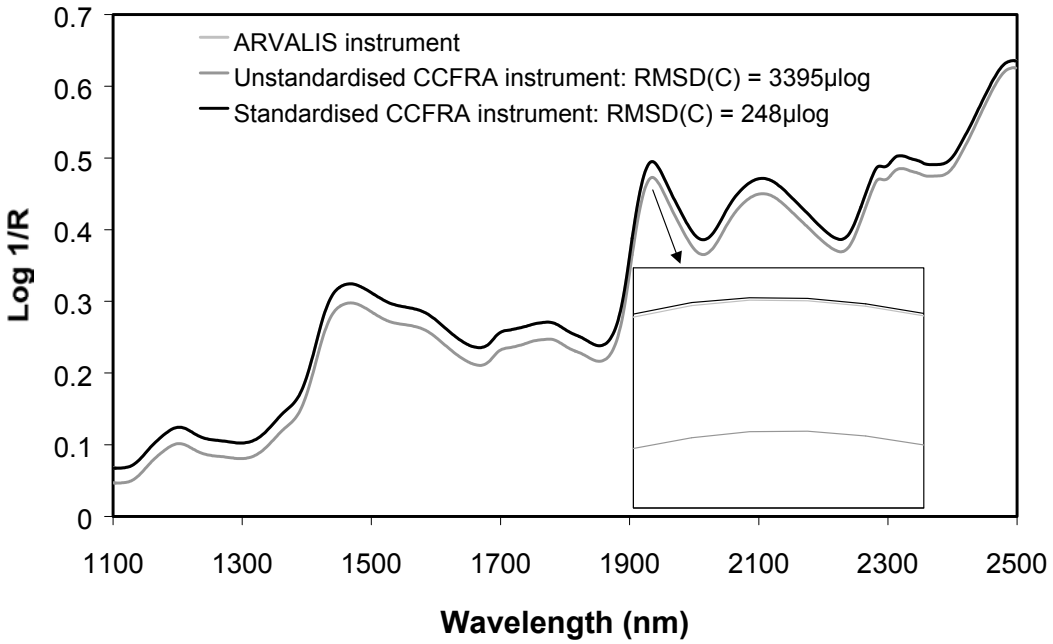
Table 11. Calibration performance for the entire flour calibration set using wavelengths typical of a filter instrument

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content as is (%)	293	9.73	1.18	0.98	0.17	0.17
Moisture content (%)	299	14.25	0.52	0.94	0.12	0.12
Water absorption (%)	302	56.29	3.65	0.84	1.44	1.46
Particle Size Index (%)	299	46.07	9.11	0.88	3.18	3.19
Hagberg Falling Number (s)	305	327.2	51.4	0.12	48.2	48.5
Damaged starch (FU)	296	24.6	8.7	0.87	3.1	3.1
Flour Colour Grade	298	-0.88	1.45	0.91	0.44	0.44
Gel protein gel mass (g)	307	8.48	1.95	0.45	1.45	1.46
Gel protein G' (Pa)	298	30.97	18.27	0.32	15.01	15.11
CBP loaf volume (ml)	290	3320.8	190.8	0.38	149.8	151.5
CBP crumb score	295	6.1	1.1	0.33	0.9	0.9
No-time dough loaf volume (ml)	211	1500.5	122.8	0.29	103.3	103.8
F1 (%)	305	13.93	1.62	0.32	1.34	1.35
F2 (%)	296	23.90	1.44	0.11	1.36	1.37
F3 (%)	300	8.18	0.68	0.10	0.64	0.65
F4 (%)	293	37.90	2.61	0.46	1.92	1.94
F5 (%)	295	15.90	1.69	0.78	0.80	0.80
F1/F2	306	0.585	0.059	0.21	0.052	0.052
(F3+F4)/F1	299	3.372	0.595	0.33	0.486	0.490

[a]Standard error of calibration
[b]Standard error of cross validation

Although real differences between samples is beneficial for calibration development, those arising from large differences between the response of different NIR instruments would not be and so the first exercise which was undertaken was an instrument standardisation routine (Shenk *et al.*, 1985). This exercise was carried out for each of the datasets studied (flour, ground wheat and whole wheat) and as an example of the importance of this exercise, the results for flour are summarised in Figure 14. This shows an example spectrum for a common sample (a sealed cell from the standardisation validation set) under three conditions of analysis: at ARVALIS, at CCFRA and at CCFRA following standardisation of the instrument at CCFRA. The last of these was so similar to that scanned at ARVALIS that it may no longer be seen on the main part of the figure due to overlap of the spectra. The inset figure shows just the region between 1930 and 1940nm and demonstrates how close the agreement is. The ideal result for such a standardisation exercise is to give differences between instruments for common samples of the same order of magnitude as those found for a single instrument when the same sample is scanned twice on it. In this case, the difference between the mean of two scans for the sealed cell assessed was 3395µlog, which is significantly higher than was obtained for duplicate packs of the same flour sample during calibration data acquisition (Table 4). Following standardisation, however, a figure of 248µlog was obtained. This is substantially better than the results for flour but the fact that this resulted from duplicate scans of a sealed cell should be borne in mind. Nevertheless, the figures demonstrate very close agreement between the spectra from the two instruments following standardisation.

Figure 14. Effect of standardisation on NIR instrument agreement



Calibrations were developed using the same NIR spectral pre-treatments that had previously been determined using the data from the 1999 harvest in the UK. The results are given in Tables 12-14. The results obtained for flour when using the combined database were all very slightly inferior to those achieved for the UK database. Although the difference in each case was small, the consistency of the direction of the differences would indicate that this was real. However, the milling procedure used at the two sites was quite different and would clearly result in flours from common wheats having different extraction rates, bran content and particle size. Rather than being a negative finding, therefore, the fact that the two databases may be combined to give results not far removed from using either of the sets alone, is very positive as it indicates improved robustness. This would give users confidence that the calibrations would be more likely to perform well with the milling procedure that they use.

The samples of ground wheat and whole wheat would be expected to give more consistent NIR spectra between the laboratories due to the more straightforward preparation regimes used. Although it is accepted that small differences between grinders will exist, these should relate only to particle size, with the relative proportions of bran and endosperm remaining constant. This is borne out in the calibration results where the results, for the combined sets were more similar to those obtained for the UK set on its own. In fact, for many of the SE-HPLC parameters, the SECV results obtained were slightly better.

The overall high level of performance for the combined datasets is important for calibrations of this type, as the subtle spectral variation which has been modelled tends to mean that the resultant calibrations will be more sensitive to differences between instruments. The results here indicate that this is a real possibility but is one that may be overcome through ensuring that instrument responses are matched and that calibrations are developed using datasets representing the range of variation (for all parameters) which may be encountered in the future. As such, the performance should give users confidence in the likely performance of the calibrations for their own situation.

5.14 Performance of samples under alternative testing regimes

Due to the differences in the breadmaking procedures and final baked products tested in the UK and France and consequent differences in quality testing procedures, it was felt that a comparison of the overall performance of the exchanged samples for each harvest year was of interest. While a sample by sample assessment for each parameter was outside the scope of the study reported, general trends with some specific results of interest may be drawn from the combined results in Tables 15-18. In all cases, the results in each laboratory have been obtained on the basis of the standard milling, testing and baking procedures in that laboratory, irrespective of the origin of the samples. For all the French-grown material, the SE-HPLC results reported were from CCFRA and the results have been ordered on the basis of performance under CBP conditions. The reverse is true of the samples supplied by CCFRA to ARVALIS. For these latter samples,

Table 12. Calibration performance for the combined flour calibration set

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content db (%)	385	11.05	1.46	>0.99	0.06	0.07
Water absorption (%)	352	55.95	3.64	0.93	0.99	1.15
F1 (%)	420	14.03	1.58	0.73	0.82	0.96
F2 (%)	426	23.78	1.64	0.54	1.12	1.22
F3 (%)	421	8.13	0.74	0.69	0.41	0.46
F4 (%)	420	38.18	3.17	0.79	1.45	1.66
F5 (%)	420	15.83	1.73	0.91	0.52	0.57
F1/F2	420	0.588	0.057	0.53	0.039	0.042
(F3+F4)/F1	420	3.369	0.626	0.80	0.279	0.334

[a]Standard error of calibration
[b]Standard error of cross validation

Table 13. Calibration performance for the combined ground wheat calibration set

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content db (%)	402	12.23	1.28	0.99	0.11	0.13
Water absorption (%)	347	55.98	3.55	0.79	1.63	1.70
F1 (%)	415	14.02	1.54	0.68	0.87	1.00
F2 (%)	418	23.83	1.57	0.53	1.07	1.24
F3 (%)	425	8.13	0.75	0.68	0.43	0.50
F4 (%)	417	38.13	3.06	0.71	1.64	1.89
F5 (%)	417	15.82	1.73	0.91	0.52	0.57
F1/F2	419	0.590	0.056	0.53	0.039	0.043
(F3+F4)/F1	413	3.339	0.585	0.70	0.321	0.372

[a]Standard error of calibration
[b]Standard error of cross validation

Table 14. Calibration performance for the combined whole wheat calibration set

Parameter	n	Mean	Standard deviation	R^2	SEC[a]	SECV[b]
Protein content db (%)	406	12.23	1.29	0.98	0.16	0.18
Water absorption (%)	356	55.95	3.64	0.71	1.97	2.17
F1 (%)	418	14.04	1.54	0.51	1.08	1.14
F2 (%)	411	23.89	1.49	0.36	1.19	1.32
F3 (%)	417	8.09	0.71	0.44	0.53	0.57
F4 (%)	416	38.12	3.00	0.66	1.75	1.92
F5 (%)	416	15.84	1.73	0.83	0.72	0.85
F1/F2	417	0.589	0.056	0.59	0.036	0.041
(F3+F4)/F1	412	3.331	0.587	0.57	0.385	0.408

[a]Standard error of calibration
[b]Standard error of cross validation

in each year, 5 of the samples supplied had been bred for biscuit making and these may be recognised by the lack of bread volume and crumb score data in the UK results section of the tables.

Overall, it may be seen that, although there were some differences for individual samples, the overall performance in each laboratory ranked in a similar fashion. It may be concluded, therefore, that generally, those wheats which perform well during CBP breadmaking will have reasonable potential for production of French bread. It is also clear that there were global differences between harvest years, with UK samples from 2000 generally giving poorer performance in both UK and French assessment. Some of this may be due to the individual samples selected although, in each case, the same approach was taken as described above, i.e. a semi-factorial selection on the basis of a range of parameters. The balance of gluten properties as assessed using SE-HPLC was also different for UK and French-grown wheats. Generally speaking, those from France had lower values, particularly for the ratio (F3+F4)/F1. This agrees with the results presented for these wheats in comparison with the main body of UK-sourced samples in the UK calibration set presented in Figure 4b.

French-grown wheats from the 1999 harvest (Table 15) gave a broad range of performance when tested under UK conditions and this range appeared to be relatively greater than for the samples when tested under a French regime. Overall the ranking of samples in terms of their test baking results were similar but there were a few exceptions to this. Samples CMS/00/142, CMS/00/146 and CMS/00/136 all gave very good performance when producing bread using the French approach, while their performance in CBP test baking was relatively poor. The SE-HPLC results for these were quite similar to those obtained for UK-grown samples and so this does not appear to explain the discrepancies observed. Only CMS/00/146 had a level of protein content which would be thought typical of UK breadmaking flours and so it may be that, for the other two samples, this has contributed to the poor performance in UK breadmaking. This was also demonstrated for sample CMS/00/145 where the protein content was again relatively low but the performance in France was better (the highest loaf volume of the set) than that in the UK.

The French-grown wheats from 2000 and 2001 (Table 16) again exhibited a broad range of properties, with some of the samples possessing the lowest values for (F3+F4)/F1 being found here. In particular, samples CMS/02/139 and CMS/02/122 appeared to have particularly strong gluten characteristics and as such were over-tough for good performance in UK breadmaking, particularly given their very low protein contents. It was interesting to note, however, that even though their performance under French testing conditions was also limited, the bread score for CMS/02/139 indicated reasonable performance, which was relatively better than that achieved in the UK. Sample CMS/02/217 was also interesting, as again, the performance under French conditions appeared better than that in the UK. The complexity of the relationships between the factors affecting baking performance was demonstrated again by the performance observed for CMS/02/216. Even with a low protein content and a lower (F3+F4)/F1 ratio than typical for UK wheats, this sample

Table 15. Relative performance of French-grown wheats from 1999 harvest

Sample Number	Variety	CCFRA					ARVALIS		
		Flour protein content db (%)	CBP loaf volume (ml)	CBP crumb score	SE-HPLC F1/F2	SE-HPLC (F3+F4)/F1	Sample Number	Loaf volume (ml)	Bread Score
CMS/00/127	Rapor	10.0	3145	6	0.64	2.96	995322	1480	196
CMS/00/134	Scipion	8.9	3161	6	0.55	2.99	995265	1480	185
CMS/00/143	Texel	9.1	3176	6	0.55	2.87	995300	1530	200
CMS/00/129	Cyrano	9.7	3242	7	0.61	2.88	995401	1570	190
CMS/00/132	Isengrain	9.3	3250	7	0.62	2.89	995333	1610	201
CMS/00/141	Sideral	9.4	3267	6	0.59	3.15	995298	1465	194
CMS/00/163	Texel	9.9	3288	6	0.55	3.01	995343	1585	204
CMS/00/131	Isengrain	10.1	3289	6	0.67	2.88	995341	1655	202
CMS/00/142	Soissons	10.6	3292	6	0.63	3.08	995299	1850	233
CMS/00/146	Courtot	14.2	3303	5	0.59	3.70	995296	1880	229
CMS/00/160	Sideral	10.7	3304	7	0.57	3.29	995351	1640	209
CMS/00/136	Soissons	11.4	3313	6	0.62	3.25	995340	1855	233
CMS/00/162	Charger	10.3	3325	7	0.56	3.09	995348	1760	224
CMS/00/159	Charger	9.6	3351	7	0.58	2.91	995309	1625	216
CMS/00/147	Shango	9.8	3367	7	0.63	2.69	995328	1770	223
CMS/00/164	Isengrain	10.4	3370	6	0.64	3.04	995338	1570	204
CMS/00/166	Shango	9.6	3374	7	0.55	3.09	995313	1730	227
CMS/00/133	Charger	9.4	3383	7	0.59	2.86	995281	1615	212
CMS/00/138	Aztec	10.5	3390	6	0.63	2.79	995302	1550	200
CMS/00/165	Soissons	10.6	3395	8	0.65	3.01	995403	1800	224
CMS/00/148	Soissons	11.4	3396	7	0.64	2.98	995334	1780	226
CMS/00/128	Aztec	9.8	3397	7	0.62	2.80	995319	1730	218
CMS/00/161	Aztec	9.2	3401	6	0.60	2.86	995279	1750	213
CMS/00/145	Soissons	10.2	3411	6	0.56	3.30	995354	1905	233
CMS/00/144	Isengrain	10.0	3422	7	0.66	2.81	995303	1725	218
CMS/00/137	Isengrain	10.8	3434	6	0.66	2.85	995140	1770	221
CMS/00/135	Shango	10.2	3445	6	0.58	3.07	995350	1830	230
CMS/00/130	Sideral	10.3	3466	7	0.63	2.83	995304	1680	210
CMS/00/140	Recital	10.0	3466	8	0.70	2.64	995339	1805	225
CMS/00/139	Galibier	13.7	3579	8	0.70	2.63	995139	-	-

Table 16. Relative performance of French-grown wheats from 2000 and 2001 harvests

Year	Sample Number	Variety	CCFRA					ARVALIS		
			Flour protein content db (%)	CBP loaf volume (ml)	CBP crumb score	SE-HPLC F1/F2	SE-HPLC (F3+F4)/F1	Sample Number	Loaf volume (ml)	Bread Score
2000	CMS/02/215	Baltimor	9.5	2866	4	0.64	2.61	012572	1365	190
2001	CMS/02/139	Autan	8.6	2929	3	0.70	2.51	010912	1565	197
2001	CMS/02/145	Boston	11.2	3021	4	0.65	3.67	010744	1295	176
2001	CMS/02/122	Isengrain	7.9	3041	4	0.74	2.28	010518	1355	179
2001	CMS/02/130	Cyrano	9.5	3140	5	0.63	2.80	010552	1410	207
2001	CMS/02/135	Sponsor	8.3	3159	3	0.59	2.72	010597	1438	176
2001	CMS/02/127	Isengrain	9.2	3168	4	0.71	2.42	010536	1352	189
2001	CMS/02/140	Nogent	-	3168	5	0.70	2.35	010918	1398	193
2001	CMS/02/119	Cezanne	9.6	3203	5	0.66	2.58	010474	1460	208
2001	CMS/02/132	Isengrain	10.3	3232	4	0.70	2.51	010567	1530	197
2001	CMS/02/125	Scipion	9.6	3252	5	0.67	2.46	010527	1380	193
2001	CMS/02/146	Autan	11.7	3274	7	0.66	2.86	010917	1568	214
2001	CMS/02/136	Baltimor	9.9	3296	4	0.63	2.57	010601	1500	200
2001	CMS/02/134	Courtot	14.0	3314	6	0.61	3.25	010589	1315	163
2000	CMS/02/218	Valoris	9.2	3319	4	0.65	3.15	012510	1565	201
2001	CMS/02/144	pParador	10.7	3322	6	0.60	2.86	011010	1528	210
2001	CMS/02/143	Balance	9.9	3326	5	0.59	3.26	011004	1480	202
2000	CMS/02/217	Aztec	10.7	3331	6	0.63	3.15	012513	1740	224
2000	CMS/02/219	Altria	9.2	3338	3	0.62	2.78	012512	1520	195
2001	CMS/02/128	Charger	9.8	3362	5	0.60	2.79	010549	1570	219
2001	CMS/02/131	Soissons	10.9	3369	7	0.69	2.64	010554	1570	217
2001	CMS/02/138	Shango	9.8	3372	5	0.59	2.84	010604	1772	229
2001	CMS/02/121	Apache	-	3374	5	0.64	2.41	010515	1560	216
2001	CMS/02/137	Sponsor	9.1	3390	5	0.60	2.85	010602	1585	197
2001	CMS/02/133	Shango	9.8	3395	5	0.60	2.86	010568	1622	222
2001	CMS/02/129	Aztec	9.8	3409	6	0.64	2.80	010551	1560	194
2001	CMS/02/124	Apache	10.1	3467	6	0.67	2.46	010523	1650	227
2001	CMS/02/116	Apache	-	3475	5	0.60	2.75	010431	1675	220
2001	CMS/02/118	Soissons	10.2	3497	7	0.66	2.64	010434	1655	216
2001	CMS/02/126	Aztec	10.5	3497	6	0.64	2.66	010530	1768	226
2001	CMS/02/120	Aztec	9.5	3557	6	0.65	2.59	010482	1660	209
2001	CMS/02/117	Galibier	14.1	3598	8	0.71	2.73	010432	-	-
2000	CMS/02/216	Texel	10.3	3727	8	0.62	2.71	012507	1750	217
2001	CMS/02/123	Soissons	-	-	-	-	-	010519	1475	205
2001	CMS/02/142	Caphorn	9.7	-	-	0.70	2.53	010760	1515	206

Table 17. Relative performance of UK-grown wheats from 2000 harvest

CCFRA				ARVALIS				
Sample Number	Flour protein content db (%)	CBP loaf volume (ml)	CBP crumb score	Sample Number	Loaf volume (ml)	Bread Score	SE-HPLC F1/F2	SE-HPLC (F3+F4)/F1
CM/53204/7126	10.0	-	-	01 2374	1162	117	0.48	4.04
CM/53204/7215	11.7	3361	7	01 2385	1185	150	0.64	3.16
CM/53204/7208	13.5	3307	7	01 2382	1212	116	0.62	3.56
CM/53204/7213	12.4	3360	7	01 2384	1212	128	0.58	4.07
CM/53204/7102	9.3	-	-	01 2373	1215	113	0.49	4.43
CM/53204/7245	9.4	3147	6	01 2387	1215	132	0.67	3.20
CM/53204/7161	12.6	2907	5	01 2379	1225	124	0.53	3.79
CM/53204/7210	13.5	3187	6	01 2383	1230	117	0.58	3.67
CM/53181/5094	11.2	3345	7	01 2392	1250	138	0.53	3.55
CM/53204/7202	9.7	-	-	01 2381	1270	148	0.50	4.08
CM/53204/7147	10.4	3272	7	01 2375	1328	169	0.54	3.70
CM/53204/7019	10.4	2991	4	01 2365	1332	170	0.60	3.42
CM/53204/7149	11.0	3461	7	01 2377	1358	192	0.55	3.39
CM/53204/7217	12.4	3313	7	01 2386	1368	180	0.57	3.25
CM/53204/7157	11.3	3251	6	01 2378	1380	173	0.56	3.82
CM/53181/5033	9.6	-	-	01 2388	1402	180	0.56	3.55
CM/53204/7095	10.3	3250	6	01 2371	1408	192	0.60	3.49
CM/53204/7081	10.6	3117	5	01 2369	1428	202	0.69	2.99
CM/53181/5068	13.2	3462	7	01 2389	1440	184	0.59	3.56
CM/53204/7033	8.7	-	-	01 2366	1450	193	0.52	3.37
CM/53204/7096	10.0	2906	6	01 2372	1450	192	0.70	3.02
CM/53204/7148	11.8	3226	6	01 2376	1478	176	0.52	3.88
CM/53181/5071	12.3	3404	7	01 2390	1515	183	0.63	3.15
CM/53204/7002	10.2	3135	6	01 2364	1532	204	0.62	2.97
CM/53204/7080	10.9	3325	6	01 2368	1550	200	0.61	4.48
CM/53204/7092	9.8	3203	5	01 2370	1560	193	0.58	4.98
CM/53204/7079	12.0	3426	8	01 2367	1610	204	0.52	3.52
CM/53204/7000	11.2	3295	7	01 2363	1632	215	0.50	3.46
CM/53204/7167	12.8	3154	6	01 2380	1635	208	0.61	3.21
CM/53181/5074	12.6	3587	8	01 2391	1665	207	0.54	3.39

Table 18. Relative performance of UK-grown wheats from 2001 harvest

CCFRA				ARVALIS				
Sample Number	Flour protein content db (%)	CBP loaf volume (ml)	CBP crumb score	Sample Number	Loaf volume (ml)	Bread Score	SE-HPLC F1/F2	SE-HPLC (F3+F4)/F1
CM/61055/5184	11.7	3084	5	013028	1310	130	0.49	4.15
CM/61055/5199	11.3	2998	5	013051	1310	162	0.57	4.78
CM/61055/5132	11.0	-	-	013031	1315	150	0.50	3.89
CM/61055/5183	11.1	3172	6	013048	1315	159	0.64	2.94
CM/61055/5181	12.8	3223	6	013045	1322	142	0.56	4.93
CM/61055/5194	11.8	3240	6	013042	1360	179	0.52	3.54
CM/61055/5082	11.9	3462	6	013034	1400	187	0.60	4.35
CM/61070/7058	11.4	3218	4	013027	1405	162	0.48	4.50
CM/61055/5060	11.6	3428	8	013029	1425	202	0.57	3.10
CM/61055/5022	10.3	-	-	013046	1430	151	0.57	4.35
CM/61070/7068	11.6	-	-	013039	1472	192	0.49	3.94
CM/61055/5180	11.9	3378	8	013054	1492	208	0.55	3.43
CM/61055/5144	11.0	3417	5	013055	1518	216	0.56	3.42
CM/61055/5228	10.1	-	-	013049	1532	209	0.61	3.37
CM/61055/5080	10.8	3290	7	013053	1610	209	0.66	2.83
CM/61055/5081	11.3	3506	7	013035	1620	216	0.60	3.99
CM/61055/5136	11.3	3411	6	013032	1620	212	0.63	3.03
CM/61055/5157	10.1	3159	4	013047	1630	206	0.62	4.11
CM/61055/5068	11.4	3555	8	013037	1662	217	0.58	3.32
CM/61070/7015	11.6	-	-	013033	1672	215	0.53	3.59
CM/61055/5057	11.7	3532	8	013052	1692	223	0.55	3.19
CM/61055/5150	10.0	3320	6	013041	1708	216	0.57	3.18
CM/61055/5013	11.6	3429	6	013038	1742	228	0.59	3.36
CM/61055/5197	10.7	3300	7	013030	1792	227	0.61	2.98
CM/61055/5079	11.6	3459	8	013044	1795	234	0.65	3.16
CM/61055/5088	11.1	3533	7	013043	1838	227	0.64	3.22
CM/61070/7002	12.7	3629	8	013036	1862	226	0.63	4.59
CM/61055/5134	12.1	3699	7	013050	1870	227	0.61	4.33
CM/61055/5087	12.5	3557	8	013040	1872	239	0.60	3.39
CM/61055/5058	13.6	3681	8	013026	1930	235	0.69	4.41

demonstrated excellent performance in UK breadmaking tests. It also performed well in the French tests, although was not the best overall in this case.

The importance of a balance of gluten characteristics and protein content was demonstrated for sample CM/53204/7096 from the 2000 UK harvest (Table 17). Here the combination of low protein content with strong gluten characteristics resulted in poor performance in CBP even though the results in the French assessment were in the top third for this sample set. The results in France generally, however, were poorer than for French-grown wheats and this also needs to be taken into account. A more extreme example of this difference in performance, however, was CM/52304/7167, which gave good performance under French testing conditions but which gave very low CBP loaf volumes. It was also interesting to note the performance for CM/53204/7033, which was grown as a biscuit wheat, it had low protein content and F1/F2 value but still managed to produce a breadmaking score of 193, better than many breadmaking wheats, under French testing conditions.

The SE-HPLC results from the UK 2001 harvest were a good example of the differences observed between French and UK-grown wheats with generally higher values of (F3+F4)/F1 being observed (Table 18). However, the results for some of these samples were very good when assessed under French breadmaking conditions, in many cases being equivalent to those obtained for French-grown wheat. The main differences between performance in French and UK testing regimes were highlighted for samples CM/61055/5082 and CM/61055/5060, where the performance in CBP was better than that when using the French method, and for CM/61055/5157, where the reverse was the case. The first two had slightly lower F1/F2 values than would be expected for French-grown wheats and higher protein contents, and thus the balance of properties appears better suited to UK processes. Sample CM/61055/5157 probably performed relatively poorly in CBP due to the lower protein content of this flour, even though the SE-HPLC results were reasonably typical of other breadmaking flours grown in the UK. The lower protein content generally appeared to be less of a problem for breadmaking using the French method and so it is not surprising that better performance under these conditions resulted.

6 Implications for levy payers and further work

Assessment of wheat functionality is one of the most important requirements of any testing regime employed by those who develop varieties, grow wheats and process grain. It is for this reason that NIR is so widely-used throughout the world for the rapid analysis of wheat protein content. While an important determinant of baking potential, it has long been recognised that the functionality of that protein is arguably the most important characteristic of any sample of wheat. The rapid analysis of these properties has, therefore, long been viewed as a ‘holy grail’ for researchers.

New NIR calibrations have been developed within this study which demonstrate good potential for use in the rapid assessment of wheat protein quality. The performance of these calibrations is generally better for flour than for ground wheat and is poorest for whole wheat. Nevertheless, useful information which could be used as a basis for screening-type assessment may be derived from each of these materials. The calibrations have been developed for use with research grade NIR instruments, which are used only in some sections of the grain industry. However, the cost of such instrumentation is now much closer to that of the instruments which are more typically found in grain receival areas and the additional information which may be derived from their use would make their purchase feasible for many users. In addition, the availability of such instrumentation at a few cereal testing laboratories means that decisions about the quality of new and established varieties of wheat may be ascertained using the calibrations developed here.

Three main areas for further development of the work presented in this report may be identified. The first of these is the supply of the calibrations to users. With this in mind, an agreement covering the details of such exploitation has been entered into by HGCA, ARVALIS and CCFRA. Under the terms of this, the calibrations which have been developed may be offered to users on a commercial basis. While the performance of the calibrations for filter instruments has been shown to be disappointing, that for scanning instruments is at a level where they may be offered for use with confidence in their likely performance. To continue to develop the calibrations and to take account of new wheat varieties as they become available, it would be prudent to develop a suitable follow-up programme which would encompass NIR scanning and, at least, SE-HPLC analysis of representative samples from forthcoming harvest years. While suitable funding for such work has yet to be identified, it is recommended that calibration development be incorporated into the annual wheat quality assessment in France and the UK, provided suitable arrangements with those supplying samples may be found.

Finally, an extension of the work to allow an assessment of wheat quality pre-harvest would represent the ultimate embodiment of the work carried out to date. CCFRA are HGCA are involved in a LINK project (Project LK0927, ‘Managing late N applications to meet wheat protein market requirements using pre-harvest near infrared (NIR) sensing’), the primary aim of which is the development of NIR calibrations for assessment of the likely protein content at maturity of wheat when assessed at an early stage of development. Given the poorer performance of protein quality calibrations obtained in this study for mature whole wheat when compared with flour, it would appear that the assessment of immature wheat may be too challenging an application for NIR. Nevertheless, previous work (Bhandari, 2000) indicates that there is potential under suitable sample processing conditions for NIR calibrations for protein quality of immature wheat to be developed. Further work will, therefore, be carried out within the current LINK project to ascertain whether early prediction of wheat protein quality as well as content is feasible.

7 Summary of key results and conclusions

- Traditional flour quality tests give useful information but do not allow baking potential to be predicted.
- Size-exclusion high performance liquid chromatography (SE-HPLC) may be used to characterise UK and French wheats in terms of their protein quality.
- UK-grown wheats tended to have higher levels of gliadin material than samples from France.
- Wheats having different end-uses may be segregated in respect of their protein characteristics using SE-HPLC data.
- NIR calibrations for protein quality and test baking characteristics may be developed which appear to have reasonable potential for rapid assessment of sample characteristics.
- Calibrations for wheat flour gave better performance than those for ground wheat which, in turn, gave better performance than those for whole wheat.
- The calibrations developed for a range of protein quality characteristics generally were not reliant on protein content, rather responding to other protein characteristics.
- Good performance was obtained for measures of protein quality when using a separate validation set while calibrations for test baking performance gave relatively poor results.
- The use of calibrations as a direct measure of test baking performance may not be recommended due to poor predictive ability.
- The spectral regions selected for use in calibrations for protein content and quality prediction were similar although their relative importance varied with parameter assessed.
- Calibrations developed using wavelengths typical of commercially available filter instruments gave unacceptably poor performance for prediction of protein quality.
- Calibrations developed using a combined dataset of samples analysed in France as well as in the UK gave more robust calibrations, indicating improved potential for general use.
- Samples giving good performance in UK or French baking evaluation generally tended to give good performance in the opposite procedure.
- The performance of flour in French baking tests was less reliant on protein content than that used for theChorleywood Bread Process (CBP) in the UK.

8 Acknowledgements

Thanks are due to Christine Bar-L'Helgouac'h, Brigitte Mahaut, Martine Giraud and Cécile Massin and the team at ARVALIS for their training in the Profilblé® method and for supply and testing of wheat and flour samples.

Thanks are due to the British Society of Plant Breeders Ltd for use of National List data and samples.

Thanks are also due to the commercial organisations who supplied samples for inclusion in the calibration database through the project.

The author is indebted to Mr Doug Smith and Dr Dhan Bhandari for the SE-HPLC analysis performed throughout the study.

The contribution of colleagues in the Cereals & Milling and Baking & Cereal Processing Departments at CCFRA is gratefully acknowledged.

Data from the Recommended List were abstracted from the Home-Grown Cereals Authority Recommended List DataBase that can be consulted at http://www.hgca.com.

9 References

Andersson, R., Hämäläinen, M. and Åman, P. (1994). Predictive modelling of the bread-making performance and dough properties of wheat. *Journal of Cereal Science*, **20**, 129-138.

Barnes, R.J., Dhanoa, M.S. and Lister, S.J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, **43**(5), 772-777.

Barnes, R.J., Dhanoa, M.S. and Lister, S.J. (1993). Correction to the description of Standard Normal Variate (SNV) and De-Trend (DT) Transformations. *Journal of Near Infrared Spectroscopy*, **1**, 186-186.

Belton, P.S. (1999). On the elasticity of wheat gluten. *Journal of Cereal Science*, **29**, 103-107.

Bhandari, D.G. (2000). The early prediction of breadmaking quality of grain and its improvement through targeted late application of nitrogen fertilisers. HGCA Report No. 219.

Bushuk, W., Briggs, K.G. and Shebeski, L.H. (1969). Protein quantity and quality as factors in the evaluation of bread wheats. *Canadian Journal of Plant Science*, **49**, 113-122.

CCFRA (2002). Guideline No. 3, Manual of Methods for Wheat and Flour Testing.

Chen, C.H. and Bushuk, W. (1970). Nature of proteins in *Triticale* and its parental species. I. Solubility characteristics and amino acid composition of endosperm proteins. *Canadian Journal of Plant Science*, **50**, 9-14.

Dachkevitch, T. and Autran, J-C (1989). Prediction of baking quality of bread wheats in breeding programs by size-exclusion high-performance liquid chromatography. *Cereal Chemistry*, **66**, 448-456.

Dardenne, P. and Biston, R. (1990). Standardisation procedure and NIR instrument network. In: *Proceedings of the 3rd International Conference on Near Infrared Spectroscopy*, 1990 (R. Biston and N. Bartiaux-Thill, eds). Agricultural Research Centre Publishing, Gembloux, pp 655-662.

Dardenne, P., Biston, R. and Sinnaeve, G. (1992). Calibration transferability across NIR instruments. In: *Near Infra-Red Spectroscopy. Bridging the Gap between Data Analysis and NIR Applications* (K.I. Hildrum, T. Isaksson, T. Naes and A. Tandberg, eds.). Ellis Horwood, New York, pp 453-458.

Delwiche, S.R. (1993). Measurement of single kernel wheat hardness using near-infrared transmittance. *Transactions of the ASAE*, **36**, 1431-1437.

Delwiche, S.R. and Weaver, G. (1994). Bread quality of wheat flour by near-infrared spectrophotometry: feasibility of modelling. *Journal of Food Science*, **59**, 410-415.

Delwiche, S. R., Graybosch, R. A., Peterson, C. J. (1998). Predicting protein composition, biochemical properties, and dough-handling properties of Hard Red Winter wheat flour by near-infrared reflectance. *Cereal Chemistry*, **75**, 412-416.

Ewart, J.A.D. (1982). Baking quality and gluten amide level. *Journal of Food Technology*, **17**, 365-372.

Farrand, E.A. (1969). Starch damage and alpha-amylase as bases for mathematical models relating to flour water absorption. *Cereal Chemistry*, **46**, 103-116.

Fearn, T. (1999). Hidden variables and partial correlations. *NIR News*, **10**, 8-9.

Fearn, T., Eddison, C., Withey, R. and Cowe, I.A. (1996). A method for wavelength standardisation in filter instruments. *Journal of Near Infrared Spectroscopy*, **4**, 111-118.

Finney, K.F. and Barmore, M.A. (1948). Loaf volume and protein content of hard winter and spring wheats. *Cereal Chemistry*, **25**, 291-312.

Graveland, A., Bongers, P. and Bosveld, P. (1979). Extraction and fractionation of wheat flour proteins. *Journal of the Science of Food and Agriculture*, **30**, 71-84.

Graveland, A., Bosveld, P., Lichtendonk, W.J., Marseille, J.P., Moonen, J.H.E. and Scheepstra, A. (1985). A model for the molecular structure of the glutenins from wheat flour. *Journal of Cereal Science*, **3**, 1-16.

Hareland, G.A. (1994). Evaluation of flour particle size distribution by laser diffraction, sieve analysis and near-infrared reflectance spectroscopy. *Journal of Cereal Science*, **21**, 183-190.

Hart, H.V. (1976). Protein determination in (home-grown) wheat by the Infra Alyzer. *FMBRA Bulletin*, 1976, No.6, 192-204.

International Association for Cereal Science and Technology (ICC) (1986). Procedure for Near Infrared (NIR) Reflectance Analysis of Ground Wheat and Milled Wheat Products. Recommendation No. 202.

MacRitchie, F. (1979). A relation between gluten protein amide content and baking performance of wheat flours. *Journal of Food Technology*, **14**, 595-601.

Martens, H. and Naes, T. (1989). Multivariate calibration. John Wiley & Sons, Chichester.

Millar, S.J. and Scotter, C.N.G. (2002). Calibration 2000 – a new, more effective approach to providing long term robust NIR calibrations. CCFRA Report CMS/REP/36752/1.

Millar, S.J., Alava, J.M. and Salmon, S.E. (1999). Using near infrared spectroscopy to assess dough development during mixing. In: *Proceedings of the 9th International Conference of Near Infrared Spectroscopy, Verona.* A.M.C. Davis and R. Giangiacomo (eds). pp 829-834.

Morel, M-H, Dehlon, P., Autran, J.C., Leygue, J.P and Bar-L'Helgouac'h, C. (2000). Effects of temperature, sonication time, and power settings on size distribution and extractability of total wheat flour proteins as determined by size-exclusion high-performance liquid chromatography. *Cereal Chemistry*, **77**, 685-691.

Oliver, G. and Pritchard, P.E. (1993). Rheology of the gel protein fraction of wheat flour. In: *Food Colloids and Polymers; Stability and Mechanical Properties.* Royal Society of Chemistry, Cambridge, pp 255-259.

Orth, R.A. and Bushuk, W. (1972). A comparative study of proteins of wheats of diverse baking quality. *Cereal Chemistry*, **49**, 268-275.

Osborne, T.B. (1907). The proteins of the wheat kernel. Carnegie Institute, Washington. Publication No. 84.

Osborne, B.G. (1983). Investigation of the performance of an improved calibration for the determination of protein in UK home-grown wheat by near infrared reflectance analysis. *Journal of the Science of Food and Agriculture,* **34**, 1441-1443.

Osborne, B.G. (1984). Investigations into the use of near infrared reflectance spectroscopy for the quality assessment of wheat with respect to its potential for bread baking. *Journal of the Science of Food and Agriculture,* **35**, 106-110.

Osborne, B.G. and Fearn, T. (1983). Collaborative evaluation of near infrared reflectance analysis for the determination of protein, moisture and hardness in wheat. *Journal of the Science of Food and Agriculture,* **34**, 1011-1017.

Osborne, B.G., Douglas, S., Fearn, T. and Willis, K.H. (1982a). The development of universal calibrations for measurement of protein and moisture in UK home-grown wheat by near-infrared reflectance analysis. *Journal of the Science of Food and Agriculture,* **33**, 736-740.

Osborne, B.G., Douglas, S. and Fearn, T. (1982b). The application of near infrared reflectance analysis to rapid flour testing. *Journal of Food Technology*, **17**, 355-363.

Osborne, B.G., Fearn, T. and Hindle, P.H. (1993). Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. Longman Scientific & Technical, Harlow.

Pawlinsky, T. and Williams, P. (1998). Prediction of wheat bread-baking functionality in whole kernels, using near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, **6**, 121-127.

Panford, J.A. (1987). Application of near-infrared reflectance spectroscopy in North America. In: *Near-Infrared Technology in the Agricultural and Food Industries*. Phil Williams and Karl Norris, eds., American Association of Cereal Chemists, St. Paul, pp 201-211.

Payne, P.I., Nightingale, M.A., Krattiger, A.F. and Holt, L.M. (1987). The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *Journal of the Science of Food and Agriculture*, **40**, 51-65.

Salmon, S.E. (2001). Protocol for Quality Assessment of Wheat Varieties in Recommended List Trials. Campden & Chorleywood Food Research Association.

Schofield, J.D. (1986). Flour proteins: structure and functionality in baked products. In: *Chemistry and Physics of Baking*. J.M.V. Blanshard, P.J. Frazier and T, Galliard (eds). Royal Society of Chemistry, London.

Shenk, J.S. and Westerhaus, M.O. (1995). Comparison of standardisation techniques. In: *Near Infrared Spectroscopy: The Future Waves. Proceedings of the 7th International Conference on Near Infrared Spectroscopy,* Montréal, Canada, 6-11 August 1995, (A.M.C. Davies and P. Williams, eds), pp 112-115.

Shenk, J.S., Westerhaus, M.O. and Templeton Jr., W.C. (1985). Calibration transfer between near infrared reflectance spectrophotometers. *Crop Science*, **25**, 159-161.

Shenk, J.S., Workman, J.J. Jr and Westerhaus, M.O. (1992). Application of NIR spectroscopy to agriculture products. In: *Handbook of Near Infrared Analysis* (D.A. Burns and E.W. Ciurczak, eds). Marcel Dekker Inc., New York, pp 383-431.

Shewry, P.R., Popineau, Y., Lafiandra, D. and Belton, P. (2001). Wheat glutenin subunits and dough elasticity: findings of the EUROWHEAT project. *Trends in Food Science and Technology,* **11,** 433-441.

Singh, H. and MacRitchie, F. (2001). Use of sonication to probe wheat gluten structure. *Cereal Chemistry*, **78**, 526-529.

Southan, M. and MacRitchie, F. (1999). Molecular weight distribution of wheat proteins. *Cereal Chemistry*, **76**, 827-836.

Tilley, K.A., Benjamin, R.E., Bagorogoza, Moses Okot-Kotber, B., Prakash, O. and Kwen, H. (2001). Tyrosine crosslinks: molecular basis of gluten structure and function. *Journal of Agricultural and Food Chemistry*, **49**, 2627-2632.

Tillmann, P., Reinhardt, T-C and Paul, C. (2000). Networking of near infrared spectroscopy instruments for rapeseed analysis: a comparison of different procedures. *Journal of Near Infrared Spectrsocopy*, **8**, 101-107.

Wesley, I.J., Skerritt, J.H., Bekes, F., Cornish, G.B and Osborne, B.G. (1998a). Near infrared analysis of a Halberd × Cranbrook doubled haploid sample set. In: *Cereals 98, Proceedings of the 48th Australian Cereal Chemistry Conference, Cairns* (L. O'Brien, A.B. Blakeney, A.S. Ross and C.W. Wrigley eds). Royal Australian Chemical Institute, Melbourne, pp 443-446.

Wesley, I.J., Larsen, N., Osborne, B.G. and Skerritt, J.H. (1998b). Non-invasive monitoring of dough mixing by near infrared spectroscopy. *Journal of Cereal Science*, **27**, 61-69.

Wesley, I. J., Uthayakumaran, R. S., Anderssen, G. B., Cornish, G. B., Bekes, F., Osborne, B. G. and Skerritt, J. H. (1999). A curve-fitting approach to the near infrared reflectance measurement of wheat flour proteins which influence dough quality. *Journal of Near Infrared Spectroscopy*, **7**, 229-240.

Williams, P.C. (1975). Application of near infrared reflectance spectroscopy to the analysis of cereal grains and oilseeds. *Cereal Chemistry*, **52**, 561-576.

Williams, P.C. (1987). Variables affecting near-infrared reflectance spectroscopic analysis. In: *Near-Infrared Technology in the Agricultural and Food Industries*. Phil Williams and Karl Norris, eds., American Association of Cereal Chemists, St. Paul, pp 143-167.

Williams, P. C., Sobering, D. C., Pawlinsky, T., Salazar, D., Klassen, B. (2000). Determination of sedimentation volumes on Canadian bread-wheat whole-meals and sieved flours. 85th Annual AACC Meeting, Kansas City, 5-9 November, 2000.