Reverse spectroscopy: bypassing the need for NIR calibration in a barley breeding program

Alison M. Kelly[1]*, Glen P. Fox[2] and Jerry Franckowiak[3]

[1]Crop and Food Science, DEEDI, Leslie Research Centre, Toowoomba, Queensland, Australia
[2]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Australia
[3]Crop and Food Science, DEEDI, Hermitage Research Station, Warwick, Queensland, Australia
*Corresponding author: Alison.Kelly@deedi.qld.gov.au

Introduction

The science of near infrared (NIR) spectroscopy is used in agricultural research to explore the quality of a grain sample. The pioneering application of NIR technology in agriculture was for the prediction of moisture and protein content of grain.[1,2] This application involved a calibration process during which a large set of samples were collected and analysed by both a chemical method and a spectral scan of the grain in order to derive a model to predict measured protein (or moisture) as a linear combination of wavelengths in the spectral data. Subsequent samples could then be scanned and the content of the protein or water predicted via the linear model. This approach removed the need for the slow, expensive destructive process of physical extraction of that compound for all grain samples.

Following the early success of determining the content of simple compounds such as protein and water, a natural extension was to predict more complex traits arising from changes in the chemical and structural composition of grain to produce an end-product. For example, barley germinates to produce malt and then malt is fermented to produce wort for the production of beer. This sequence of processing introduces the notion of levels of degree of removal from the original grain sample which undergoes both physical and chemical changes to produce malt. This intermediate state is in turn processed into an end-product of beer. When grain is scanned using NIR spectroscopy, the degree of removal and hence difficulty in prediction of an end-product trait increases with the additional number of stages in the process.[3]

Adoption of NIR predictions for more complex traits has been limited in scientific applications, as the accuracy and reliability of the method is solely dependent on the calibration process. Variability in the reference method is the major constraint on the calibration process. For example, when measuring the extract of malt from barley, the prediction of malt content from NIR spectroscopy can only be as accurate and reliable as the measurements made from a physical micro-malting process. As multiple parameters contribute to end product quality, the traditional approach has been to predict individual traits for each quality parameter from individual calibration equations e.g. malt extract, diastatic power, fermentability and soluble nitrogen for beer. The errors in calibration for each trait are then compounded when these independent predictions are combined in a multivariate analysis to predict overall grain quality. Furthermore, the calibration equations for different traits are often based on common wavelengths as the traits are inter-related and influenced by the same underlying chemistry.

The protein-starch matrix that binds molecules together in the grain kernel regulates the ability of grain to undergo physical and chemical changes. When there is strong chemical coupling between the molecules in this matrix, the distinct vibrational behaviour of the independent components as measured by NIR spectroscopy is distorted. An analysis of complete spectra bypasses the need for calibration, and removes both the calibration error introduced and compounded through the prediction of each individual trait, as well as the potential loss of the dependence structure between wavelengths.

A sample of grain sourced from a structured plant breeding program has the ability to produce a high quality end-product due to a range of factors. Accurate quantification of the variability due to these factors maximises genetic gain when selecting for superior traits. The two major factors contributing to grain quality are the inherent genetic potential within a plant and the influence of environment on plant growth and grain development. Advanced statistical methods are required to partition the information contained in the spectral response of grain into components due to genotype (G) and environment (E) arising from underlying moisture and fertility gradients in the field.

Linear mixed models are routinely employed in cereal breeding programs in Australia for the analysis of agronomic traits of yield[4] and grain size.[5] These methods are based on spatial analysis to account for field trend[6] and an appropriate underlying variance structure for genotype by environment interaction.[7] A statistical model is proposed for modelling variation in the whole spectra and partitioning this into effects due to genotype and environment using a linear mixed model. The paper demonstrates that this approach can

Reference paper as:
Kelly A.M., Fox G.P., and Franckowiak J. (2012).Reverse spectroscopy: bypassing the need for NIR calibration in a barley breeding program, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 7-10.

be used for selection of genotypes in a cereal breeding program, removing the need for robust calibration equations for a broad range of quality parameters. The potential of the analysis for genetic selection is demonstrated using grain samples from an early generation field trial in an Australian barley breeding programme.

Materials and Methods

The dataset adopted for the case study of the whole spectra analysis method is taken from Stage 1 of testing in the northern node of the Australian barley breeding program in 2009. The field trial consisted of 1104 plots arranged in a rectangular array of 24 columns by 46 rows and was designed as a partially duplicated spatial design with two replicates of 189 genotypes and single plots of the remaining 726 genotypes.

Grain was harvested from the field trial and a subsample of plot yield was scanned through an NIRSystems 6500 using WinISI software. Spectra were exported as raw absorbance bands (log 1/R) for wavelengths ranging from 1100 to 2500 nm in steps of 2 nm, forming a total of 700 wavelengths. Existing calibration equations were used to predict various traits, but primarily hot water malt extract, the most important trait for the malting industry. Chemical analysis was undertaken on a small subset of samples measuring friability, malt extract, soluble and total nitrogen, wort β-glucan, diastatic power and fermentability as per the European Brewery Convention methods. The correlation between predicted malt extract from calibration and measured malt extract from the malting process in the laboratory was also calculated.

A combined analysis was conducted for the whole spectra of grain samples with wavelengths ranging from 700 to 2500 nm in steps of 2 nm. The data were analysed in two stages due to the dimensionality of the problem. Firstly, a principal components analysis was conducted to reduce the spectra to five components, as these components accounted for 99.65% of the variation in the spectral data. The principal component scores were then analysed using a multivariate linear mixed model approach fitting terms for genotype and accounting for the spatial correlation at the field plot level.[6] Variance heterogeneity between the five components was estimated at both the genotype and the field plot level. Variance parameter estimates were obtained using residual maximum likelihood (REML)[8] and best linear unbiased predictions (BLUPs)[9] of scores were obtained for the genotypes. The program ASReml[10] was used to fit the linear mixed model and to form predictions.

Results

The first three components of the PCA model explained 98.8% of the total variation in the spectra. This variation was partitioned into genetic and non-genetic effects via the linear mixed model which explained 58.3% and 40.5% of the variance, respectively. Spatial variation was evident and spatial correlations were estimated as 0.48 and 0.28 in the field column and row dimensions respectively. Key wavelengths with maximum loadings were determined and an example of the loadings for the wavelength at 1460 nm has been plotted on the biplot in Figure 1. Wet chemistry measurements from the subset of samples were correlated with the components and regressions for three of these quality traits are superimposed onto the biplot. The superposition of traits indicates an area of the biplot where genotypes possess favourable quality characteristics. Comparisons can be made relative to the standard genotypes that have been marked on the biplot.

Discussion and Conclusion

The factors contributing to measurement of grain quality, and indeed all agronomic traits, are the inherent potential within a plant due to genetic make-up, and the influence of environment on plant growth and grain development at the micro-environmental level due to field trends resulting from moisture and fertility gradients. The analysis proposed demonstrates that genetic and micro-environmental variation can be modelled from whole spectral data. Furthermore, the ratio of genetic to field plot variation is substantial (59:41) and is of an order of magnitude useful for making genetic gain.

The calibration for malt extract in this study was reasonably reliable ($r = 0.62$), but not accurate enough for selection of genotypes. In general, calibration equations for prediction of traits in early generation material suffer continually from the introduction of new genetic sources which may lie outside the range of the set of samples used for calibration, going against recommendations for a robust calibration.[11]

A sample of grain sourced from a structured plant breeding program has the ability to produce a high quality end-product due to a range of factors, and it is the accurate quantification and understanding of these contributing factors which maximises genetic gain when selecting for superior traits. Analysis of whole spectral data preserves the correlation between wavelengths and the inherent relationship between traits

Reference paper as:
Kelly A.M., Fox G.P., and Franckowiak J. (2012).Reverse spectroscopy: bypassing the need for NIR calibration in a barley breeding program, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 7-10.

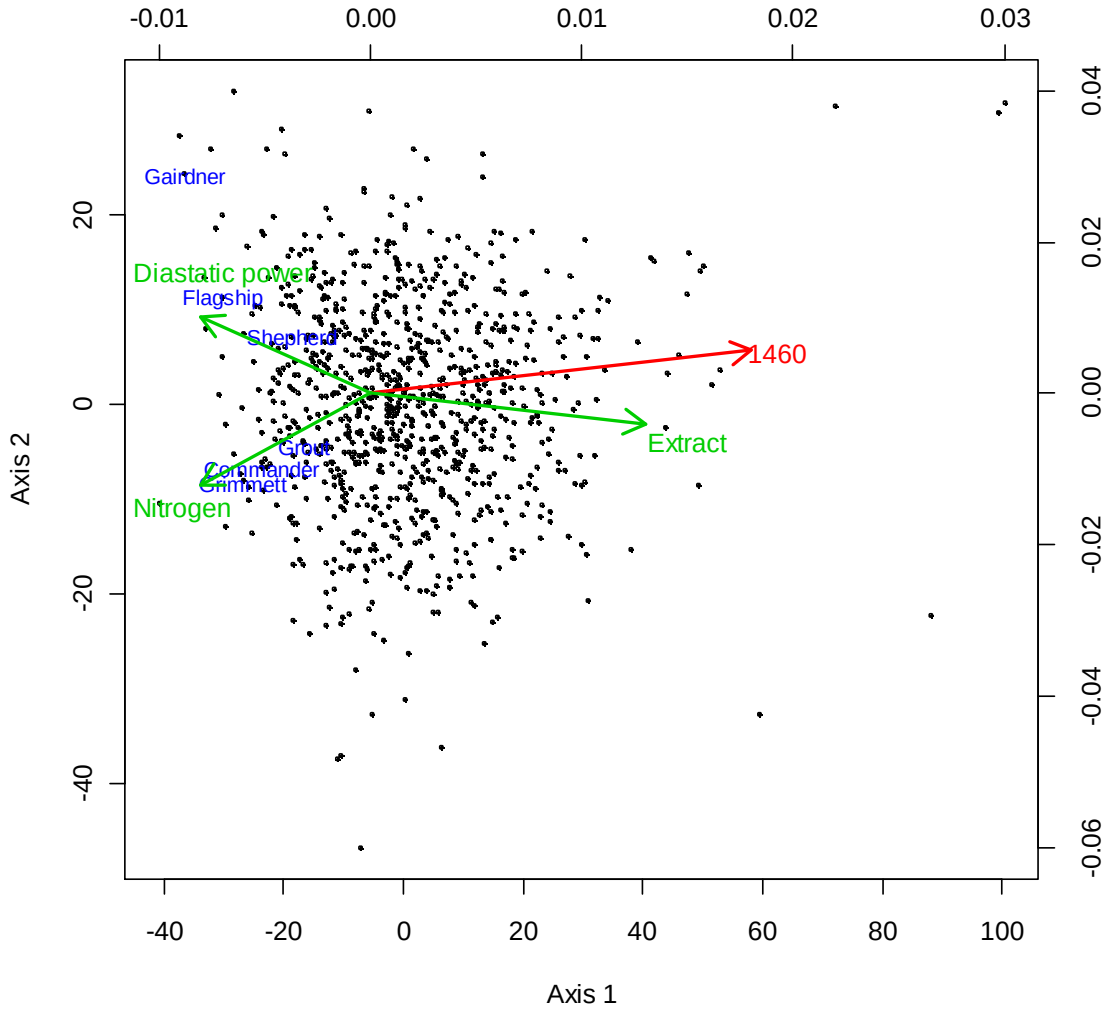

Figure 1. A biplot of genotype scores with six labelled genotypes, and loading vector for one wavelength at 1460 nm. Vectors of wet chemistry measurements for the three traits of extract, nitrogen and diastatic power were superimposed onto the biplot via regression.

measuring grain functionality. While calibration may be avoided, interpretation of patterns from whole spectral analysis still relies on laboratory measurement of key traits. These traits together with discriminating wavelengths must be superimposed on the pattern of response from whole spectral analysis to indicate favourable subsets of genotypes for further testing. Cluster analysis could also be used to group genotypes with superior mean performance of malt quality, as well as other traits.

The large number of genotypes tested in the early stages of a breeding program, 915 in this study, prevents detailed and expensive quality testing on all samples. NIR can be used as a high throughput, low cost screening tool to indicate subsets of test lines with a favourable combination of quality characteristics, relative to the control genotypes. These test lines can then be taken for further detailed testing with some confidence by using the NIR screening as an indicator of underlying grain chemistry.

Advanced statistical methods are required to partition the information contained in the spectral response of grain into components due to genotype and environment and the interaction between these, while simultaneously adjusting for the differences due to changing micro-environmental effects. While a two-stage exploratory approach was used in this study, extensions to a fully efficient one-stage analysis within a linear mixed model framework following the method of Smith *et al.*[7] are under current research.

References

1. K.H. Norris, *Agric. Eng.* **45**, 370-372 (1964).
2. P.C. Williams, *Cereal Chem.* **52,** 561-576 (1975).
3. R. Anderssen, B.G. Osborne and I. Wesley, The pragmatics of NIR calibration-predict methodology and software, Technical report, GRDC (2001).

Reference paper as:
Kelly A.M., Fox G.P., and Franckowiak J. (2012).Reverse spectroscopy: bypassing the need for NIR calibration in a barley breeding program, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 7-10.

4. A.M. Kelly, A.B. Smith, J.A. Eccleston and B.R. Cullis, *Crop Sci.* **47,** 1063-1070 (2007).
5. G.P. Fox, A.M. Kelly, D.M.E. Poulsen, P.A. Inkerman and R.J. Henry, *J. Cereal Sci.* **43,** 198-208 (2006).
6. A.R. Gilmour, B.R. Cullis and A.P. Verbyla, *J. Agric. Biol. Environ. Stat.* **2,** 269-273 (1997).
7. A.B. Smith, B.R. Cullis and R. Thompson, *Biometrics* **57,**1138-1147 (2001).
8. H.D. Patterson, and R. Thompson. *Biometrika* **31,**100-109 (1971).
9. G.K. Robinson, *Stat. Sci.* **6,** 15-51 (1991).
10. A.R. Gilmour, B.R. Cullis, B.J. Gogel, S.J. Welham and R. Thompson. *ASReml, user guide. Release 2.0*. VSN Inernational Ltd., Hemel Hempstead, UK (2005).
11. P. Dardenne, G. Sinnaeve and V. Baeten, *J. Near Infrared Spectrosc.* **8,** 229-237 (2000).

Reference paper as:
Kelly A.M., Fox G.P., and Franckowiak J. (2012).Reverse spectroscopy: bypassing the need for NIR calibration in a barley breeding program, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 7-10.