Detection of grain protein in standing wheat crops using hyperspectral sensing

Glenn J Fitzgerald[1,2], Cassandra Walker[1,2], Sahand Assadzadeh[1], Eileen Perry[3], Alex Clancy[1], Joe Panozzo[1,2]

[1] Agriculture Victoria, 110 Natimuk Road, Horsham, VIC 3400; Email: glenn.fitzgerald@agriculture.vic.gov.au

[2] Centre for Agricultural Innovation, School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, VIC 3010

[3] Agriculture Victoria, 1 Taylor Street, Epsom, VIC, 3551

Abstract
Being able to measure and map grain protein variation across paddocks before harvest, in the standing crop using sensors, would provide growers with timely information to plan for selective harvest. This could potentially lead to obtaining better grain prices and map N removal, increasing profits and improving year to year N management. Previous research using narrowband hyperspectral data to predict grain protein has focused on single sites. In this study, data were collated from a range of experiments that were conducted with wheat in paddocks and on research farms across several years and locations. Previous studies have shown that narrowband information can predict grain protein, but it is unclear whether this can be generalised such that predictive relationships are robust across locations and years. In the experiments used for this analysis, physical samples were collected from under the sensor at harvest and wheat analysed for protein and moisture content. Sensor readings were calibrated to reflectance in the field and were collected from 0-22 days prior to harvest, depending on the experiment. Data analysis showed that the field data were significantly noisier than lab-based measurement of grain protein but there are clear spectral signals for grain protein based on different calibration and validation approaches. Although on-header protein sensors are commercially available, data from satellites and unmanned aerial vehicles could expand growers' ability to utilize sensing technologies enabling planning for harvest operations to improve profit and N management.

Keywords
Spectral analysis, partial least squares, selective harvest, N management, principal components

Introduction
Near infrared spectroscopy is an established method for measuring cereal grain protein in the lab and uses contiguous spectral information (Williams and Norris, 2001). In a field context, this type of data is referred to as 'hyperspectral'. In the lab, grain can be measured, producing a 'pure' spectral signal of the target. In a standing crop in the field, spectra represent 'mixed' signals as they combine multiple canopy components within the sensor field of view, such as spikes, stems, soil, etc. Thus, the question of whether a protein signal can be resolved from the mixed signal will determine if this approach is viable for field detection and mapping of grain protein.

It has been demonstrated that there is a protein signal for wheat within the mixed field spectra (Apan et al., 2006; Fitzgerald et al., 2007) but studies have been performed on single experiments. Whether the calibrations developed for one data set can be transferred to other data sets collected under real-world relationships (with variation due to different sky and field conditions) is unknown and is an important determinant for broader use in mapping grain protein before harvest. Grain protein concentration can change due to many factors, including soil type, soil water status, fertilizer input, during the grain development and by crop type; and varies across the landscape. In wheat, grain protein can be a determinant of price and market-grade classification, with premiums paid for grain with higher protein contents. Thus, being able to map grain protein before harvest could allow growers to create protein zones and plan for harvest to achieve superior prices.

There are on-header commercial sensors available for farmers to map grain protein at harvest (Clancy, 2019). These are useful for mapping protein for on-the-go segregation of grain. Having information

before harvest would provide additional information potentially for segregating areas within paddock (zoning) for growers to make more informed decisions. This project analysed ground-based hyperspectral data collected from field and paddock experiments sown to wheat conducted across eight site-years that included grain protein analyses.

Methods

Field data used in analysis

Hyperspectral point-based sensor data were collected from each site-year (Table 1) using an Analytical Spectral Devices Fieldspec 3. From 2005-2018 data were collected with the same instrument from field experiments and paddocks in Victoria. In all cases, the spectral data were collected by positioning the sensor over wheat canopies and calibrating to reflectance using a Spectralon panel during collection. Although clear sky conditions were targeted, given the range of sites and years, sky conditions would be expected to be variable, introducing noise into the data, in addition to the various canopy components within the sensor field of view. Each spectral footprint was about 0.7 m diameter at the top of canopy. Across all data sets, biomass cuts were sampled from under the sensor at harvest, grain was separated from the wheat heads, dried at 40° C, weighed and protein measured using the Dumas combustion method.

The eight data sets used in this analysis (Table 1) were collected from wheat across a wide range of growing conditions in semi-arid rainfed and irrigated environments in research plots and in-paddock. Spectral data were collected between the day of harvest and up to 22 days before harvest.

Table 1. Data sets used in the analysis. Name, year, location, dates spectra collected, and grain harvested. The last column indicates the diversity of the data: data collected from research plots or grower paddocks; whether conditions were rainfed (R), irrigated (I) or both; or collected from experiments with different times of sowing (TOS).

Data set & year	Location	Spectra collected	Grain harvested	Notes (R=Rainfed, I=Irrigated, TOS = Time of Sowing)
FISE 2004	Maricopa, AZ (USA)	18 May 2004	26 May 2004	Plots (I)
ORL 2004	Horsham, VIC	3 Dec 2004	15 Dec 2004	Plots (R & I)
ORL 2005	Horsham, VIC	21 Nov 2005	13 Dec 2005	Plots (R & I)
ORL 2006	Horsham, VIC	8 Dec 2006	8 Dec 2006	Plots (R & I)
AGF 2007	Horsham, VIC	27 Dec 2007	12 Dec 2007	Plots, TOS1 (R & I)
		9 Jan 2008	20 Dec - 2 Jan 2008	Plots, TOS2 (R & I)
AGF 2008	Horsham, VIC	8 Dec 2008	8 & 15 Dec 2008	Plots, TOS1 & 2 (R & I)
MRS 2008	Walpeup, VIC	10 Nov 2008	10 & 25 Nov 2008	Plots, TOS1 & 2 (R & I)
FF 2018	Kaniva, VIC	28 Nov 2018	28 Nov 2018	Paddock (R)

Data analysis

The spectral data were processed using the following steps: 1) The full range of spectral data (350-2500 nm) were subsampled to the 750-1750 nm range (Figure 1a), which is typical for laboratory protein analysis, 2) the spectra were converted to log (1/R), (where R=reflectance value) and a multiplicative scatter correction was applied (Figure 1b, 3) calibrations were developed with the pre-processed spectra using a partial least square (PLS) regression to predict grain %N (dry weight basis). The number of PLS components (Table 2, N_{comp}) used in each of the calibrations was found by optimizing the validation RMSE values. Two model validation strategies used to determine the ability to predict grain N content were: 1) leave-year-out (one trial at a time selected as validation set and remaining used for calibration set), 2) random subsampling (25% of data selected for validation). Two criteria the data are required to meet for these analyses are: a) validation reference values must lie within the distribution of the calibration set and b) spectra must be similar, as determined by a principal component analysis (PCA) score plot (not shown).

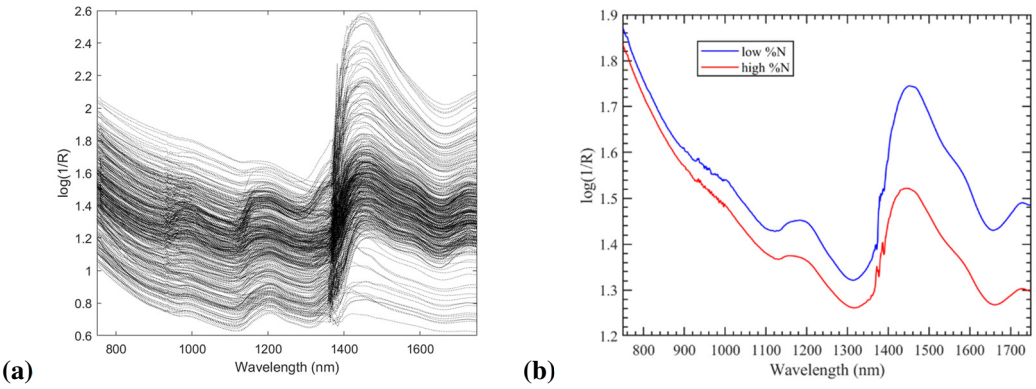

Figure 1. (a) Hyperspectral data collected from field samples with 750-1750 nm wavelength range used in analysis. (b) Example of high and low grain %N spectra, 1/R log transformed (R=reflectance).

Results and Discussion

For the leave-year-out approach, R^2 values for the validation sets ranged from 0.05 to 0.88 and RMSE 0.13 to 0.68 (Table 2 and Figure 2a – with a low and high R^2 data set). When validation samples were selected randomly with 25% of the global data used for prediction, R^2 values ranged from 0.60 to 0.72 and RMSE 0.25 to 0.29 for five different, randomly selected data sets (Table 2 and Figure 2b with example 1:1 line fits). The random sampling predicted the fit better because it met the criteria above (a and b) for the data, while for the leave-year-out approach, points aligned within the calibration data cloud, but not every year contained the full range of possible variation for grain %N.

In operation on-farm, global calibrations would be developed across sites and years and a leave-year-out method could be used to predict grain protein for a given paddock in the current year. As seen in Table 2, for some site-years this could be very successful for prediction but not in others. There was an overall better fit to the data using the random approach. Operationally, this approach could be implemented by including a small subset of the data from the new year in the calibration set to predict protein for the remainder of the samples in the set.

A more accurate approach to validation than the global approach would be through development of individual calibrations using PCA to group different historical data sets, representing different sets of conditions (which might be a combination of environment, cultivar, lighting conditions, etc.). During spectral collection before harvest, a few small samples of developing grain could be collected and based on PCA of the grain spectra, the appropriate calibration curve selected.

Alternatively, a global calibration as presented here may be sufficient to delineate two or three zones in a paddock that could be useful for managing operations. An on-harvester protein sensor could then be used to measure %N (or protein) content more accurately within each zone at harvest.

Table 2. Paired calibration and validation (cal/val) results for grain %N predictions using two approaches (leave-year-out and random 25%). N = number of data points in validation, RMSE = root mean square error, N_{comp} = No. of PLS components used to build the validation relationship.

	Leave-year-out					Random sampling			
Data set (see Table 1)	N	R^2	RMSE	N_{comp}	Set	N	R^2	RMSE	N_{comp}
FISE 2004	374/36	0.63/0.50	0.30/0.32	8	1	307/103	0.80/0.72	0.22/0.25	19
ORL 2004	364/46	0.81/0.45	0.20/0.51	19	2	307/103	0.72/0.66	0.26/0.27	14
ORL 2005	363/47	0.55/0.88	0.32/0.22	7	3	307/103	0.66/0.62	0.29/0.28	9
ORL 2006	389/21	0.78/0.80	0.23/0.23	19	4	307/103	0.81/0.60	0.22/0.29	21
AGF 2007	362/48	0.03/0.18	0.51/0.13	1	5	307/103	0.82/0.68	0.21/0.26	24
AGF 2008	268/142	0.55/0.37	0.35/0.37	6					
MRS 2008	394/16	0.70/0.20	0.27/0.45	10					
FF 2018	356/54	0.52/0.05	0.31/0.68	5					

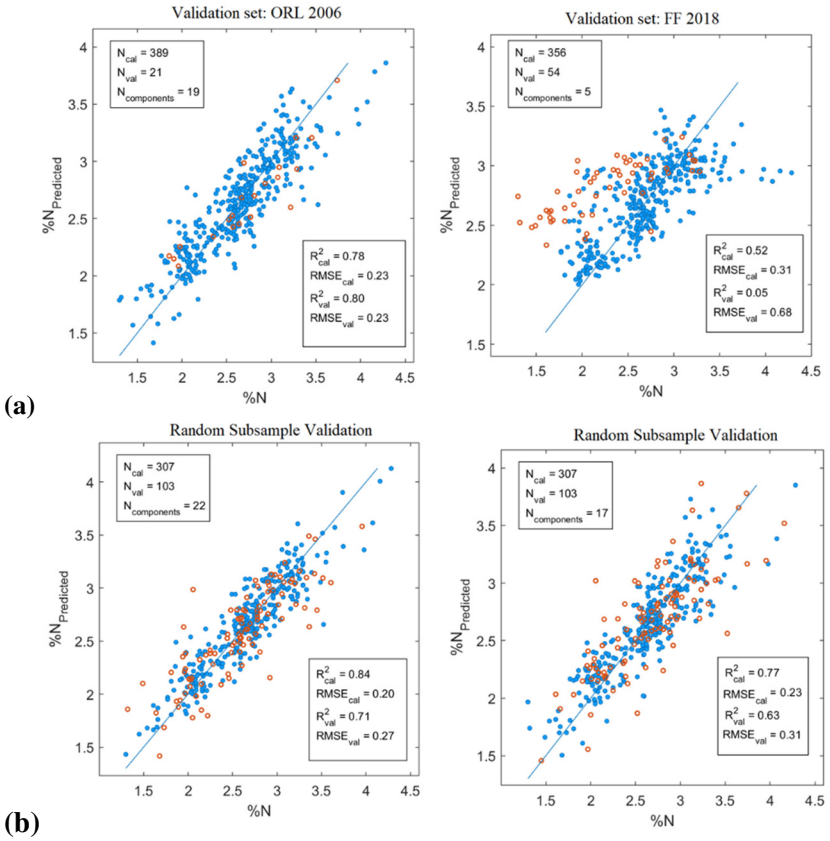

Figure 2. Examples of line fits of predictions for a 'good' season and a 'bad' fit for (a) leave-year-out, and the two examples of predicted fit for (b) random sample. Line fits shown along 1:1 line with regression statistics for calibration (cal) and validation (val) data.

Conclusion

Using hyperspectral data to quantify grain %N (or protein) just prior to harvest could assist growers make more informed decisions to harvest grain to meet market quality grades, leading to higher prices received. Although predictive equations based on field-collected spectral data is much noisier than lab-based predictions of grain protein, it appears that, despite this noise, reliable calibrations can be established allowing estimation of grain protein and delineation of manageable zones for harvest.

As hyperspectral sensors become more readily available on ground and aerial-based platforms, this type of data will be collected more routinely. Establishing calibrated relationships is fundamental to realising delivery of data to growers that can be easily managed and interpreted. Further analyses may explore targeted selection of wavelengths specific to protein detection, which could inform multispectral approaches to protein detection and mapping, useable on aerial and satellite platforms.

References

Apan A, Kelly R, Phinn S, Strong W, Lester D, Butler D, Robson A (2006). Predicting Grain Protein Content in Wheat Using Hyperspectral Sensing of In-season Crop Canopies and Partial Least Squares Regression. International J Geoinformatics, vol. 2, 93-108.

Fitzgerald GJ, Golebiowski T, Black CK (2007). Remote sensing to predict nitrogen content of wheat grain before harvest. Proceedings Royal Australian Chemical Institute (RACI) Conference, 6-10 Aug 2007, Melbourne, Australia.

Clancy P (2019). Finding the sweet spot in nitrogen fertilization by measuring protein with an on combine NIR analyser. In: Grains: A Global Food Resource, Proceedings of the 69th Australasian Grain Science Conference, p 106-108, 27th – 29th August 2019, Melbourne, Victoria.

Williams PC, Norris K (2001). Near-infrared technology in the agricultural and food industries, 2nd Ed. Am. Assoc. Cereal Chem., St. Paul, MN.