Estimation of Chemical Composition of Quinoa (*Chenopodium quinoa* Willd.) by Near-Infrared Transmission Spectroscopy

Christian Encina-Zelada[1,2,3], Vasco Cadavez[1], Jorge Pereda[2], Luz Gómez-Pando[4], José A. Teixeira[3] & Ursula Gonzales-Barron[1]

[1]CIMO Mountain Research Centre, School of Agriculture, Polytechnic Institute of Braganza, Portugal. E-mail: **cencina@ipb.pt**

[2]Faculty of Food Industries, National Agricultural University La Molina, Lima, Peru

[3]Department of Biological Engineering, School of Engineering, University of Minho, Portugal

[4]Faculty of Agronomy, National Agricultural University La Molina, Lima, Peru

Introduction

Despite being a less well-known plant, there has been increasing interest in quinoa (*Chenopodium quinoa* Willd.) for the past 15 years owing to its perceived superior nutritional quality in comparison with other grains. Near infrared transmission (NIT) spectroscopy can presently provide rapid and low-cost whole kernel analysis of starch, moisture, protein, and oil percentages in cereals. **Thus, the objective of this study was to develop partial least square (PLS) models to accurately quantify protein and fat contents by NIT spectroscopy.**

Proximal and NIT analyses

- Quinoa crops harvested in Peru were utilised. Quinoa grains samples amounted to 73 accessions which were of orange, beige, black and yellow colour.

- Near-infrared transmission (NIT) spectra were acquired by placing the grains directly in an Infratec 1241 grain analyzer (Module Foss Tecator), using 60-mm quartz cuvettes, and scanning the region 850-1050 nm. The spectra were recorded at scanning step intervals of 2 nm to give 100 data points per sample. All raw spectral data were then mean-centered and linked to the chemical analyses data. To minimise the effect of changes in the baseline, the raw spectra were firstly pre-processed using the following filters separately: **multiplicative scatter correction (MSC), first and second derivatives using the Savitzky-Golay method (SG1, SG2), standard normal variate (SNV)** and **detrending (DT).** In addition, combinations of filters with MSC were also tested: **SG1+MSC, SG2+MSC, SNV+MSC** and **DT+MSC**.

Statistical analyses

Separate PLS analyses were carried out using protein and fat contents as dependent variables. For a specified number of PLS components, the cross-validation was set to randomly remove 10 samples at once (prediction set), and estimate the root mean square error of prediction (RMSEP) and the coefficient of determination (R^2) for the plot between the values predicted from the NIT model and the chemical analyses observations. For each of the nine pre-processing filters, RMSEP and R^2 statistics were obtained for a number of PLS components ranging from 10 to 20. Thus, the optimal numbers of Partial Least Squares (PLS) factors were deduced for every pre-processing filter.

Results

Applying the first and then the second derivative emphasized the peaks below and above the baseline, providing the best resolution for the expected signals (Fig. 1).

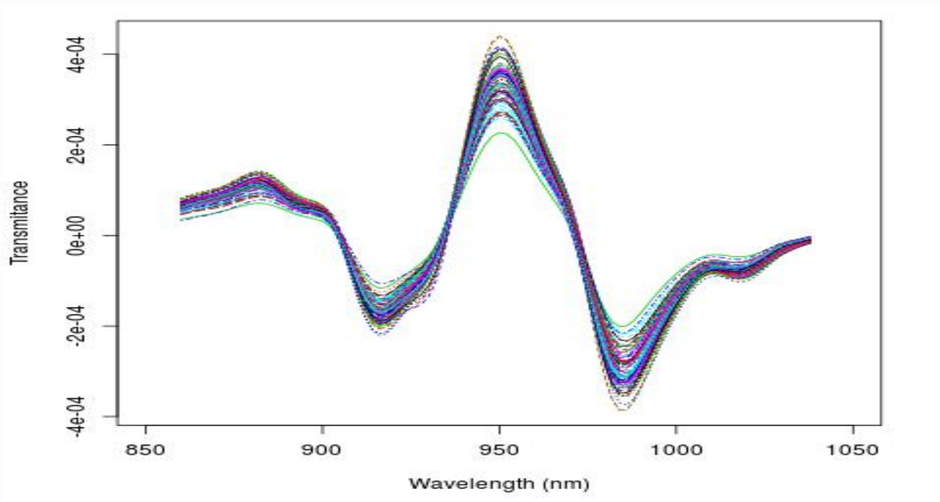

Figure 1: Savitzky-Golay Second Derivative Transformed-Spectral Data From Quinoa Grains

In all cases, it was observed that as more PLS components were retained, higher R^2 were achieved. However, as the prediction ability of the model cannot be based solely on R^2 values, the RMSEP values were primarily evaluated. RMSEP was found to exhibit a different behaviour: in most cases RMSEP steadily decreased until attaining a minimum value (at an optimal number of PLS components), at which point they increased at a faster pace as more PLS components were retained (Fig. 2).

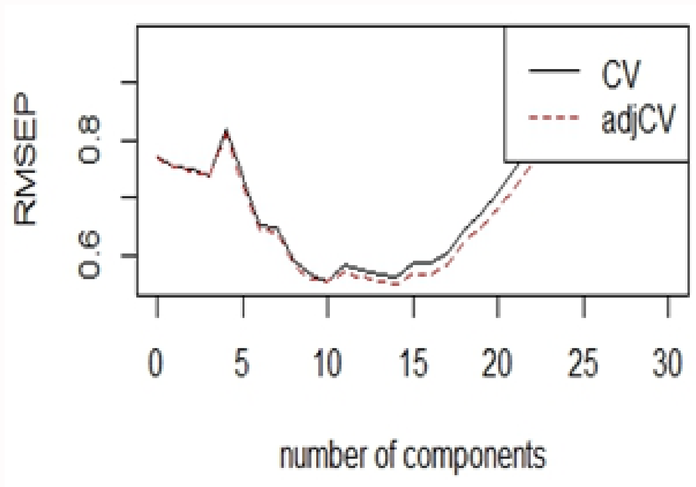

Figure 2: RMSEP Values as Affected by Number of PLS Components, Extracted from Spectra Pre-processed by Detrending and Multiplicative Scatter Correction for Protein

For fat, R^2 ranges from 0.588 to 0.815, while RMSEP ranges from 0.423 to 0.607%. For protein, R^2 ranges from 0.681 to 0.897, while RMSEP ranges from 0.523 to 0.627% (Table 1). Subsequently, to find an optimal number of PLS components, RMSEP was kept to a minimum while R^2 was targeted to a reasonable value (>0.70). Thus, in terms of model predictability for fat content, in general SG2, SNV and DT filters led to more accurate results than the other filters. Specifically, extracting 18 PLS components from the SG2 pre-processed spectra (RMSEP=0.432; R^2=0.728) produced comparable results to extracting 16 PLS components from the SNV-processed spectra (RMSEP=0.438; R^2=0.754) and the DT-processed spectra (RMSEP=0.451; R^2=0.786).

In the case of protein estimation, greater predictability of the models was generally achieved by the SG2 and DT filters (Table 1). Specifically, extracting 16 PLS components from the SG2+MSC pre-processed spectra produced statistics (RMSEP=0.598; R^2=0.859) that were comparable to the extraction of 14 components from the DT-processed spectra (RMSEP=0.545;R^2=0.833) and the DT+MSC-processed spectra (RMSEP=0.597; R^2=0.862).

Table 1: RMSEP adj (%) Statistic and R^2 (in brackets) Obtained from Cross-Validation for Protein Quantification Using Individual and a Combination of Pre-Processing Filters

Filter applied	Number of PLS Components			
	10	**12**	**14**	**16**
SG1 +MSC	0.538 (0.695)	0.542 (0.716)	0.627 (0.754)	0.626 (0.813)
SG2	0.596 (0.693)	0.614 (0.757)	0.610 (0.777)	0.603 (0.796)
SG2 +MSC	0.523 (0.737)	0.556 (0.78)	0.606 (0.819)	**0.598 (0.859)**
DT	0.546 (0.695)	0.587 (0.7905)	**0.545 (0.833)**	0.619 (0.875)
DT +MSC	0.547 (0.681)	0.582 (0.816)	**0.597 (0.862)**	0.617 (0.897)

Conclusions

- Unlike the models for fat estimations, the application of MSC in conjunction with other filters (SG2, DT) consistently enhanced the prediction capacity of the PLS models.
- The PLS models estimating fat yielded overall greater accuracy than those estimating protein content, as indicated by their lower RMSEP values.
- **Best predictions for fat** were obtained by extracting 16-18 partial least square (PLS) components from **SNV**, **DT** and **SG2**-treated spectra (mean square error of prediction, RMSEP=0.432-0.451; R^2=0.728-0.786) while **best predictions for protein** were attained by extracting 14-16 PLS components from **DT**, **SG2+MSC** and **DT+MSC**-treated spectra (RMSEP=0.545-598; R^2=0.833-0.862).

Dr. Gonzales-Barron wishes to acknowledge the financial support provided by the Portuguese Foundation for Food Science and Technology through the award of a five-year Investigator Fellowship (IF) in the mode of Development Grants (IF/00570).